

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 13:43:01 ; Search time 3357 Seconds  
(without alignments)  
6029.549 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctatccacct.....taataaacagttaaaagctg 467

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : GenEmbl.\*

1: gb\_bat.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_srs.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sta.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_man.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	% Match	Query Length	DB ID	Description
1	467	100.0	494	6	AX092298 Sequence
2	467	100.0	494	6	AX696953 Sequence
3	467	100.0	495	9	AY358657 Homo sapi
4	467	100.0	2673	9	BC021104 Homo sapi
5	467	100.0	50000	6	AR211704 Sequence
6	467	100.0	151152	9	HS454M7 Human DNA
7	467	100.0	206618	9	AF195953 Homo sapi
8	454	97.2	2243	6	BD156791 Primer fo
9	454	97.2	2243	9	AK001855 Homo sapi
10	454	97.2	2243	9	AK001855 Homo sapi
11	419.2	89.8	458	6	AX332625 Sequence
12	419.2	89.8	458	6	AX332852 Sequence
13	391.2	83.8	403	6	BD154391 Primer fo
14	391.2	83.8	403	6	BD154391 Rattus no
15	151.8	32.5	171310	2	AC127934 Rattus no
16	151.8	32.5	208810	2	AC136814 Rattus no
17	129.8	27.8	3143	10	BC020015 Mus muscu
18	129.8	27.8	118604	10	AL714010 Mouse DNA
19	129.8	27.8	189982	10	AL672274 Mouse DNA
20	129.8	27.8	203805	10	AC055817 Mus muscu
21	127	27.2	598	9	H02B38F03
22	74.2	15.9	255	6	AX211242 Sequence
23	46.4	9.9	73778	6	AX344562 Sequence
24	46.4	9.9	349980	6	AX344561 Sequence
25	45	9.6	134971	2	AC116367 Oryza sat
26	45	9.6	212887	2	AC134496 Rattus no
27	45	9.6	272030	2	AC105482 Rattus no
28	44.2	9.5	71781	2	AC101159 Mus muscu
29	44.2	9.5	213417	2	AC119515 Rattus no
30	44.2	9.5	242655	2	AC107465 Rattus no
31	44.2	9.5	243369	10	AL606973 Mouse DNA
32	44	9.4	173254	2	AC146436 Pan trogl
33	43.8	9.4	903	3	AF441245 Nephila c
34	43.8	9.4	155665	2	AC133581 Mus muscu
35	43.2	9.3	119790	9	HSDJ989D7 Human DNA
36	43	9.2	59414	9	AC093112 Homo sapi
37	43	9.2	343188	2	AC114853 Rattus no
38	42.8	9.2	224334	2	AC096041 Rattus no
39	42.6	9.1	575	10	M15614 Mouse germl
40	42.6	9.1	982	10	M20136 Mouse beta-
41	42.6	9.1	2708	3	M76700 Drosophila
42	42.6	9.1	145052	10	AL603889 Mouse DNA
43	42.6	9.1	199221	2	AC128156 Rattus no
44	42.6	9.1	250611	10	AE000663 Mus muscu
45	42.6	9.1	261600	10	AC125228 Mus muscu
46	42.4	9.1	139786	10	AC124343 Mus muscu
47	42.4	9.1	155276	2	AC091992 Homo sapi
48	42.4	9.1	202733	10	AC122805 Mus muscu
49	42.4	9.1	203124	2	AC131668 Mus muscu
50	42.4	9.1	234228	2	AC125852 Rattus no
51	42.2	9.0	50783	2	AC120601 Rattus no
52	42.2	9.0	192814	9	AC007597 Homo sapi
53	42.2	9.0	214786	9	AC007610 Homo sapi
54	42.2	9.0	227164	2	AC103418 Rattus no
55	42.2	9.0	227663	2	AC097301 Rattus no
56	42.2	9.0	240105	2	AC117339 Rattus no
57	42.2	9.0	241730	2	AC103283 Rattus no
58	42.2	9.0	283782	2	AC096338 Rattus no
59	42	9.0	55116	9	AP005854 Homo sapi
60	41.8	8.9	217240	5	BX004795 Zebrafish
61	41.6	8.9	110000	2	Continuation (3 of
62	41.6	8.9	142278	9	AC005177 Homo sapi
63	41.6	8.9	241865	2	AC126834 Rattus no
64	41.6	8.9	315654	2	AC118159 Rattus no
65	41.4	8.9	1585	5	AY225852 Ctenophar

```

66 41.4 8.9 190402 2 AC114715 AC114715 Rattus no
67 41.2 8.8 6014 9 AB033028 AB033028 Homo sapi
68 41.2 8.8 61020 6 AX251545 AX251545 Sequence
69 41.2 8.8 174423 2 AC122951 AC122951 Rattus no
70 41.2 8.8 268907 2 AC107588 AC107588 Rattus no
71 41.2 8.8 309024 2 AC107574 AC107574 Rattus no
72 41 8.8 160594 2 AC122006 AC122006 Mus muscu
73 41 8.8 179645 2 AC119245 AC119245 Mus muscu
74 41 8.8 193311 2 AC118694 AC118694 Mus muscu
75 41 8.8 270357 2 AC123318 AC123318 Rattus no
76 40.8 8.7 5349 6 AX345242 AX345242 Sequence
77 40.8 8.7 46643 10 AL732512 AL732512 Mouse DNA
78 40.8 8.7 119160 2 AC135533 AC135533 Rattus no
79 40.8 8.7 125020 9 AF429315 AF429315 Homo sapi
80 40.8 8.7 244594 2 AC112470 AC112470 Rattus no
81 40.8 8.7 250361 2 AC123175 AC123175 Rattus no
82 40.8 8.7 289156 2 AC130870 AC130870 Rattus no
83 40.6 8.7 202000 2 AC118476 AC118476 Mus muscu
84 40.6 8.7 228385 2 AC145731 AC145731 Mus muscu
85 40.6 8.7 229752 10 AC102652 AC102652 Mus muscu
86 40.6 8.7 240276 2 AC097773 AC097773 Rattus no
87 40.4 8.7 961 5 AF137232 AF137232 Gadus mor
88 40.4 8.7 3218 3 DMABF X76217 D.melanog
89 40.4 8.7 3583 3 AY069823 AY069823 Drosophil
90 40.4 8.7 4004 3 DMGATAPAC Y07662 D.melanog
91 40.4 8.7 7218 6 I66494 I66494 Sequence 14
92 40.4 8.7 53427 2 AC019841 AC019841 Drosophil
93 40.4 8.7 103709 9 AC090135 AC090135 Homo sapi
94 40.4 8.7 131556 2 AC132404 AC132404 Mus muscu
95 40.4 8.7 161160 9 AC083964 AC083964 Homo sapi
96 40.4 8.7 173660 3 AC007755 AC007755 Drosophil
97 40.4 8.7 177278 3 AC007726 AC007726 Drosophil
98 40.4 8.7 186180 2 AC126037 AC126037 Mus muscu
99 40.4 8.7 186765 2 AC016992 AC016992 Homo sapi
100 40.4 8.7 190130 10 AC122808 AC122808 Mus muscu

```

## ALIGNMENTS

```

RESULT 1
AX092298 LOCUS AX092298 494 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 29 from Patent WO0116318.
ACCESSION AX092298
VERSION AX092298.1 GI:13444463
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and
Wood,W.I.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 29 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source Location/Qualifiers
1 .494
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 5.3e-113;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTTGCCTATCCACCTCCCGAGCCCTTACCTATGCTGCTGCTAAACGCTGCTG 60
Db 1 CAATGTTTGCCTATCCACCTCCCGAGCCCTTACCTATGCTGCTGCTGCTGCTGCTG 60

```

```

QY 61 CTGCTGCTGCTGCTCTTAAAGGCTCATGCTGGAGTGGGAGCTGCTGGTCCCGAGAAA 120
Db 61 CTGCTGCTGCTGCTCTTAAAGGCTCATGCTGGAGTGGGAGCTGCTGGTCCCGAGAAA 120
QY 121 GTCTCTTCTGCACTGACGCCCCCATCAGGATTTGGGCTTCTTCCCTTCCCTTCTG 180
Db 121 GTCTCTTCTGCACTGACGCCCCCATCAGGATTTGGGCTTCTTCCCTTCCCTTCTG 180
QY 181 TGTCTCTGCTCATGCGCCCTGCGCATGACCTGCAGCAAGCCAGCCCGTGGGAGGG 240
Db 181 TGTCTCTGCTCATGCGCCCTGCGCATGACCTGCAGCAAGCCAGCCCGTGGGAGGG 240
QY 241 GAGAAAGTGGGGAGGCTAAGAAAGCTGGAGATAGGAGACAGAGAGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGAGGCTAAGAAAGCTGGAGATAGGAGACAGAGAGGTAGTGGGTG 300
QY 301 GGCTAGGGGGCTGCTCTTAAAGTGGTGTATGATCTTATCTTATCTTATCTTATCAA 360
Db 301 GGCTAGGGGGCTGCTCTTAAAGTGGTGTATGATCTTATCTTATCTTATCTTATCAA 360
QY 361 AGATATTAAAGCCCTGTTTCAATTAAGAAATGTTCCCTTCCCTTCCCTTCAATGTTGTAA 420
Db 361 AGATATTAAAGCCCTGTTTCAATTAAGAAATGTTCCCTTCCCTTCCCTTCAATGTTGTAA 420
QY 421 AGATTTCTGTGTAATATGCTTTTATATAAAGTGGTGTATGATCTTATCTTATCTTATCAA 467
Db 421 AGATTTCTGTGTAATATGCTTTTATATAAAGTGGTGTATGATCTTATCTTATCTTATCAA 467
RESULT 2
AX696953 LOCUS AX696953 494 bp DNA linear PAT 02-APR-2003
DEFINITION Sequence 21 from Patent WO0078961.
ACCESSION AX696953
VERSION AX696953.1 GI:29497962
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
Ferrara,N., Stewart,T.A., Williams,P.M., Baker,K.P., Deanoyers,L.,
Eaton,D.L., Gao,W.Q., Pan,J., Botstein,D., Fong,S., Goddard,A.,
Godowski,P.J., Gurney,A.L., Smith,V., Tumas,D., Wood,W.I.,
Grimaldi,C.J., Hillan,K.J., Paoni,N.F., Roy,M.A. and Watanabe,C.K.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0078961-A 21 28-DEC-2000;
Genentech Inc. (US)
FEATURES
source Location/Qualifiers
1 .494
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 5.3e-113;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTTGCCTATCCACCTCCCGAGCCCTTACCTATGCTGCTGCTAAACGCTGCTG 60
Db 1 CAATGTTTGCCTATCCACCTCCCGAGCCCTTACCTATGCTGCTGCTGCTGCTGCTG 60
QY 61 CTGCTGCTGCTGCTCTTAAAGGCTCATGCTGGAGTGGGAGCTGCTGGTCCCGAGAAA 120
Db 61 CTGCTGCTGCTGCTCTTAAAGGCTCATGCTGGAGTGGGAGCTGCTGGTCCCGAGAAA 120
QY 121 GTCTCTTCTGCACTGACGCCCCCATCAGGATTTGGGCTTCTTCCCTTCCCTTCTG 180
Db 121 GTCTCTTCTGCACTGACGCCCCCATCAGGATTTGGGCTTCTTCCCTTCCCTTCTG 180

```

```

QY 181 TGCTCTGCTGCTCATGGCTGCGCTGACCTGACGCGCAAGCCAGCCCGCTGGGGAAGGG 240
Db 181 TGCTCTGCTGCTCATGGCTGCGCTGACCTGACGCGCAAGCCCGCTGGGGAAGGG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
QY 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 360
Db 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 360
QY 361 AGATATTAAAGGCGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 420
Db 361 AGATATTAAAGGCGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 420
QY 421 AGATTGTTCTGTGTAATATGTCTTTATAATAACAGTTAAAGCTG 467
Db 421 AGATTGTTCTGTGTAATATGTCTTTATAATAACAGTTAAAGCTG 467

RESULT 3
AY358657 495 bp mRNA linear PRI 03-OCT-2003
LOCUS Homo sapiens clone DNA56962 APELIN (UNQ471) mRNA, complete cds.
ACCESSION AY358657
VERSION AY358657.1 GI:37182435
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J.,
Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B.,
Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E.,
Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S.,
Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C.,
Schoenfeld,J., Seehagiri,S., Simmons,L., Singh,J., Smith,V.,
Stinson,J., Vagstad,A., Vandlen,R., Watanabe,C., Wleand,D., Woods,K.,
Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z.,
Goddard,A., Wood,W.I. and Godowski,P.
The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
Effort to Identify Novel Human Secreted and Transmembrane Proteins:
A Bioinformatics Assessment
Genome Res. 13 (10), 2265-2270 (2003)
JOURNAL PUBLISHED 12975309
REFERENCE 2 (bases 1 to 495)
AUTHORS Clark,H.F.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Department of Bioinformatics, Genentech,
Inc., 1 DNA Way, South San Francisco, CA 94080, USA
FEATURES
source
1. 495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DNA56962"
1. 495
/locus_tag="UNQ471"
40. 261
/locus_tag="UNQ471"
/feature="PR0831"
/codon_start=1
/product="APELIN"
/protein_id="AAQ89020.1"
/db_xref="GI:37182436"
/translators="MILLITLILLLKSLGLENGVGAQKVSSATDAPIRDNAPFP
SFCLLPHPAMTCSQAPRGEKVGDG"

Query Match 100.0%; Score 467; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 5.3e-113;

```

```

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTGCTATCCACCTCCCGCCCAAGCCCTTTACCTATGCTGCTGCTGAACGCTGCTG 60
Db 1 CAATGTTGCTATCCACCTCCCGCCCAAGCCCTTTACCTATGCTGCTGCTGAACGCTGCTG 60
QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGAGTGGGAGCTGGTGGCTGCCAGAAA 120
Db 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGAGTGGGAGCTGGTGGCTGCCAGAAA 120
QY 121 GTCTCTTCTGCCACTGACGCCCGCCCATCAGGAGTGGGCTTCTTCCCCCTTCTTCTG 180
Db 121 GTCTCTTCTGCCACTGACGCCCGCCCATCAGGAGTGGGCTTCTTCCCCCTTCTTCTG 180
QY 181 TGTCTCTGCTCATGGCTGCGCTGACCTGACGCGCAAGCCCGCTGGGGAAGGG 240
Db 181 TGTCTCTGCTCATGGCTGCGCTGACCTGACGCGCAAGCCCGCTGGGGAAGGG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTG 300
QY 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 360
Db 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 360
QY 361 AGATATTAAAGGCGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 420
Db 361 AGATATTAAAGGCGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTATACAA 420
QY 421 AGATTGTTCTGTGTAATATGTCTTTATAATAACAGTTAAAGCTG 467
Db 421 AGATTGTTCTGTGTAATATGTCTTTATAATAACAGTTAAAGCTG 467

RESULT 4
BC021104 2673 bp mRNA linear PRI 07-OCT-2003
LOCUS Homo sapiens apelin, mRNA (cDNA clone MGC:31846 IMAGE:4586949),
complete cds.
ACCESSION BC021104
VERSION BC021104.1 GI:18088893
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2673)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,D., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Donald,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalius,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL PUBLISHED 22388257
MEDLINE 12477932
REFERENCE 2 (bases 1 to 2673)
AUTHORS Strausberg,R.

```

**TITLE**  
JOURNAL

**REMARK**  
**COMMENT**

Direct Submission  
Submitted (03-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/PTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiac, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 40 Row: f Column: 18  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 21314667.

**FEATURES**  
source

1..2673  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:31846 IMAGE:4586949"  
/tissue\_type="Kidney, renal cell adenocarcinoma"  
/clone\_lib="NIH MGC 14"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
1..2673  
/gene="APLN"  
/db\_xref="LocusID:8862"  
420..788  
/db\_xref="MIM:300297"  
/codon\_start=1  
/product="apelin"  
/protein\_id="AAH21104.1"  
/db\_xref="GI:18088894"  
/db\_xref="LocusID:8862"  
/translations="MSATWCSPEGQMGQPGREVGNAGSPASPIRDPCLSEAGL  
KGPPSAHPRRLCLHLVCFSGSLTSIQLSPRTCCSHQWALFSPACFPQWRAPGCSL  
DSDRSRLTRIPVHLPPSLD"

**ORIGIN**

Query Match 100.0%; Score 467; DB 9; Length 2673;  
Best Local Similarity 100.0%; Pred. No. 5.7e-113;  
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60  
Db 2188 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 2247

QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGCTGCTGCGTCCAGAAA 120  
Db 2248 CTGCTGCTGCTGCTTAAAGCTCATGCTGGAGTGGGACTGCTGCTGCGTCCAGAAA 2307

QY 121 GTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCTTCTTTCCCTTCTCTG 180  
Db 2308 GTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCTTCTTTCCCTTCTCTG 2367

QY 181 TGTCTCTGCTCATCGGCTGCCATGACCTGAGCCAGCCAGCCCGTGGGAAGGG 240  
Db 2368 TGTCTCTGCTCATCGGCTGCCATGACCTGAGCCAGCCAGCCCGTGGGAAGGG 2427

QY 241 GAGAACTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAAGAGGGTACTGGGTG 300  
Db 2428 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAAGAGGGTACTGGGTG 2487

QY 301 GGCTAGGGGGCTGCTTATTATTAAAGTGGTGTGTTATGATCTTATATCTAATTTATCAA 360  
Db 2488 GGCTAGGGGGCTGCTTATTATTAAAGTGGTGTGTTATGATCTTATATCTAATTTATCAA 2547

QY 361 AGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 420  
Db 2548 AGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 2607

QY 421 AGATTGTTCTGTGTAATAATGTTCTTTATAATAACACAGTTAAAAAGCTG 467  
Db 2608 AGATTGTTCTGTGTAATAATGTTCTTTATAATAACACAGTTAAAAAGCTG 2654

RESULT 5  
AR211704/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

AR211704  
Sequence 4 from patent US 6399349.  
AR211704  
AR211704.1 GI:21515097  
Unknown.  
Unclassified.  
1 (bases 1 to 50000)  
Ryan, J.W. and Sprinkle, T. Joe. Curtis.  
Human aminopeptidase P gene.  
Patent: US 6399349-A 4 04-JUN-2002;  
Location/Qualifiers  
1..50000  
/organism="unknown"  
/mol\_type="unassigned DNA"

**ORIGIN**

Query Match 100.0%; Score 467; DB 6; Length 50000;  
Best Local Similarity 100.0%; Pred. No. 6.4e-113;  
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60  
Db 1518 CAATGTTGGCTATCACCTCCCGAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 1459

QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGCTGCTGCGTCCAGAAA 120  
Db 1458 CTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGCTGCTGCGTCCAGAAA 1399

QY 121 GTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCTTCTTTCCCTTCTCTG 180  
Db 1398 GTCTCTTCTGCCACTGACGCCGCCCATCAGGGATTGGGCTTCTTTCCCTTCTCTG 1339

QY 181 TGTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Db 1338 TGTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279

QY 241 GAGAAAGTGGGGATGGCTTAAAGAAAGCTGGGAGATAGGGAACAAGAGGGTAGTGGGTG 300  
Db 1278 GAGAAAGTGGGGATGGCTTAAAGAAAGCTGGGAGATAGGGAACAAGAGGGTAGTGGGTG 1219

QY 301 GGCTAGGGGGCTGCTTATTATTAAAGTGGTGTGTTATGATCTTATATCTAATTTATCAA 360  
Db 1218 GGCTAGGGGGCTGCTTATTATTAAAGTGGTGTGTTATGATCTTATATCTAATTTATCAA 1159

QY 361 AGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 420  
Db 1158 AGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 1099

QY 421 AGATTGTTCTGTGTAATAATGTTCTTTATAATAACACAGTTAAAAAGCTG 467  
Db 1098 AGATTGTTCTGTGTAATAATGTTCTTTATAATAACACAGTTAAAAAGCTG 1052



```

RESULT 6
HS454M7/c
LOCUS      151152 bp      DNA      linear      PRI 05-JUN-2003
DEFINITION Human DNA sequence from clone RP3-454M7 on chromosome Xq25-26.3,
complete sequence.
ACCESSION AL022162
VERSION    HTG.
KEYWORDS   OT22162.1 GI:3171881
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151152)
AUTHORS   Pavitt,R.
TITLE     Direct Submission
JOURNAL   Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 2, 1998 this sequence version replaced gi:2969945.
COMMENT   ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormmap This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP3-454M7 is from the library RPCI-3 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-454M7.
location/Qualifiers
1. 151152
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704M07454"
/db_xref="taxon:9606"
/chromosome="X"
/map="q25-26.3"
/clone="RP3-454M7"
/clone_lib="RPCI-3"
767. 35998
/gene="OCRL1"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"

```

## FEATURES

## source

```

/note="variant 1"
match: CDNAS: Em:M88162 Em:U57627 Em:M7A161 Em:AF040094
match: ESTs: Em:AA368192 Em:AA704671 Em:AA515789
Em:AA02623 Em:AA126320 Em:T63686 Em:AA884344 Em:N92504
Em:AA044611 Em:AA188493 Em:AA743649 Em:AA836673 Em:R67320
Em:AA100629 Em:AA085500 Em:R94403 Em:AA056506 Em:AA034375
Em:AA142870 Em:AA150871 Em:T84251 Em:AA122020 Em:AA906612
Em:F07337 Em:AA628152 Em:AA878369 Em:AA640853 Em:AA189134
Em:AA044666 Em:AA740555 Em:R18793 Em:AA844284 Em:AA904845
Em:AA042798 Em:AA122019 Em:H53971 Em:W38961 Em:AA805220
Em:AA868822 Em:AA032176 Em:AA034374 Em:T84250 Em:N56932
Em:AA169401 Em:AA188849 Em:N46002 Em:AA056392 Em:HB7857
Em:W52373 Em:AA587050 Em:AA100630 Em:T88888 Em:H27722
Em:AA186750 Em:R66483 Em:T28294"
/evidence=not_experimental
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1.2 (variant 2)"
/protein_id="CAAL18150.1"
/db_xref="GI:3171882"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLPCHRNRYFLMAFLRELLKSEYNSVNMNMIATLFTSLLRPPNLMAR
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
EIVMLGLDNLRYCLMDANVKSLINKKDIQRLIKPQNLQRTQKKAQFVDNPEIK
FIPTYKDSKTRDSSGKCRVPAWCDRIILWRGTVNQLNRYRSHMELKTSDEKZVSAL
FHIGVQDERRRYKPFEDSVIRMDRMENDFLPSELSRREFPENVKFRLOQKEFQ
ISNGQVPCFHSFIPKLNDSQYCKPWLRAEPFEGYLEPNETVDSLIVYVSKDSQTL
NSGDKIEDILVLHLDKQDYFLTISGNLYPSCFTGSLALCRMKRIRREVPTKLID
LEDSFLEKESLLQWPLDEGASERPLQPKETIWLVLVLFKACHQEDLFOFGMQ
BEIQQIICDCLDTSIPETIPGNSHVAELILEALPEPVICVELYORCLDSAYDRI
QVTSQOLRAIQLFLGLGSEED"
join(<767..877,1303..1392,2075..2195,2282..2443,
3992..4093,4621..4735,5826..5942,6041..6228,9214..9325,
10696..10805,12706..12841,18582..18692,19339..19504,
19759..19994,30444..30560,31621..31705,32328..32455,
33287..33398,33588..33998)
/gene="OCRL1"
/note="variant 1"
/codon_start=1
/evidence=not_experimental
/product="dJ454M7.1.1 (Low Oculocerebrorenal Syndrome)"
/db_xref="GI:4160528"
/translation="GCKIRVQGWIRERREIPEDEHCLKFLSAVLAQAQSQLLVP
EQDSSWYQKLDITKPKPSVSGLLGFEDNFSSMLDKKINSQNPQGIHREPPPPF
SVNKLPREKEANSKQPKVNTNRKLFVQNTQSGQREGLIKHLAKREKEYNIQTF
RFFVGTWVNGQSPDSGLPWLNDPNPDYICIGFOELDLSTFAFFYFVSKEQWS
MAVERGLHSAKAKYKVLQVLGVMLLIIFARKQCYRIIDTATVGTGIVGKMGNG
GVAVRFVHNTTFCIVNSHLAAVEDPFRNODYKDCARMSFVVPQTLFQNLIMKH
```

```

repeat_region /note="MER33 repeat: matches 1. .323 of consensus"
1178. .8315
repeat_region /note="L1PA2 repeat: matches 5000. .6146 of consensus"
11225. .11294
repeat_region /note="MER5B repeat: matches 109. .178 of consensus"
11345. .11657
repeat_region /note="AluYb8 repeat: matches 1. .310 of consensus"
11736. .11797
repeat_region /note="31 copies 2 mer tt 72% conserved"
12436. .12562
repeat_region /note="L2 repeat: matches 2579. .2710 of consensus"
13276. .13457
repeat_region /note="AluSg/x repeat: matches 129. .310 of consensus"
13679. .13873
repeat_region /note="MIR repeat: matches 5. .213 of consensus"
14132. .14206
repeat_region /note="MIR repeat: matches 163. .233 of consensus"
14207. .14413
repeat_region /note="MER8 repeat: matches 2. .239 of consensus"
14414. .14548
repeat_region /note="MIR repeat: matches 12. .163 of consensus"
14793. .15073
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
15383. .15589
repeat_region /note="MER3 repeat: matches 1. .209 of consensus"
16139. .16279
repeat_region /note="MER5B repeat: matches 29. .178 of consensus"
16317. .16466
repeat_region /note="MIR repeat: matches 52. .212 of consensus"
16667. .17000
repeat_region /note="L1PA9 repeat: matches 5829. .6163 of consensus"
17001. .17301
repeat_region /note="AluYb8 repeat: matches 1. .302 of consensus"
17302. .17642
repeat_region /note="L1PA9 repeat: matches 5491. .5829 of consensus"
18285. .18620
repeat_region /note="2 copies 168 mer 78% conserved"
18772. .19981
repeat_region /note="MIR repeat: matches 6. .242 of consensus"
18945. .19005
repeat_region /note="L2 repeat: matches 2648. .2702 of consensus"
20349. .20489
repeat_region /note="MIR repeat: matches 1. .150 of consensus"
20543. .20697
repeat_region /note="MER5A repeat: matches 22. .188 of consensus"
21496. .21830
repeat_region /note="L1MC4 repeat: matches 7477. .7849 of consensus"
22420. .22650
repeat_region /note="MER46A repeat: matches 1. .236 of consensus"
22894. .22965
repeat_region /note="3 copies 24 mer 83% conserved"
24309. .24606
repeat_region /note="AluSx repeat: matches 1. .299 of consensus"
25081. .25134
repeat_region /note="27 copies 2 mer ta 70% conserved"
25168. .25193
repeat_region /note="13 copies 2 mer tg 100% conserved"

Query Match 100.0%; Score 467; DB 9; Length 151152;
Best Local Similarity 100.0%; Pred. No. 6.7e-113;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTTGCTATCCACTCCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
Db CAATGTTTGCTATCCACTCCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 89659

QY 61 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 89658

QY 121 GTCCTCTTGCCACTGACGCCGCCCATCAGGATGCGGCTCTTTCCCTCCCTTCTG 180
Db GTCCTCTTGCCACTGACGCCGCCCATCAGGATGCGGCTCTTTCCCTCCCTTCTG 89539

```

```

QY 191 TGTCTCTGCTCATCGCCCTGCCCATGACCTGCAGCCCAAGCCACCCCGTGGGAAGG 240
Db TGTCTCTGCTCATCGCCCTGCCCATGACCTGCAGCCCAAGCCACCCCGTGGGAAGG 89479

QY 241 GAGAAAGTGGGGGATGCTAGAGAAAGCTGGGAGATAGGAAAGAGAGGCTAGTGGGTG 300
Db GAGAAAGTGGGGGATGCTAGAGAAAGCTGGGAGATAGGAAAGAGAGGCTAGTGGGTG 89419

QY 301 GGCTAGGGGGCTCCCTTATTAAAGTGGTGTGTTATGATCTTATATACTAATTATACAA 360
Db GGCTAGGGGGCTCCCTTATTAAAGTGGTGTGTTATGATCTTATATACTAATTATACAA 89359

QY 361 AGATTAAAGCCCTGCTCATTAAGAAATGTTCCCTCCCTGCTGTTCAATGTTTAA 420
Db AGATTAAAGCCCTGCTCATTAAGAAATGTTCCCTCCCTGCTGTTCAATGTTTAA 89358

QY 421 AGATTGTTCTGTAAATATGCTTTTATAATAACAGTTAAAGCTG 467
Db AGATTGTTCTGTAAATATGCTTTTATAATAACAGTTAAAGCTG 89252

RESULT 7
AF195953 206618 bp DNA linear PRI 26-MAR-2002
LOCUS Homo sapiens membrane-bound aminopeptidase P (XNPEP2) gene,
DEFINITION complete cds.
ACCESSION AF195953
VERSION AF195953.2 GI:19718557
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C.
TITLE Human membrane-bound aminopeptidase P genomic DNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 206618)
AUTHORS Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Vascular Biology Center, Medical College of
Georgia, 1120 15th Street, Augusta, GA 30912, USA
REFERENCE 3 (bases 1 to 206618)
AUTHORS Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C.
JOURNAL Direct Submission
SUBMITTED (26-MAR-2002) Vascular Biology Center, Medical College of
Georgia, 1120 15th Street, Augusta, GA 30912, USA
REMARK Sequence update by submitter
COMMENT On Mar 26, 2002 this sequence version replaced gi:11066156.
FEATURES
Location/Qualifiers
source 1..206618
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
144189..176791
/gene="XNPEP2"
join(144189..144502,147348..147421,149203..149313,
150445..150508,151478..151582,151837..151923,
152849..152995,155710..155811,156987..157068,
157392..157587,158401..158490,159714..159823,
160536..160613,161726..161797,164422..164482,
165754..165823,166414..166518,167248..167307,
167936..168012,172845..172934,173533..176791)
/product="membrane-bound aminopeptidase P"
144189..144453
/gene="XNPEP2"
join(144454..144502,147348..147421,149203..149313,
150445..150508,151478..151582,151837..151923,
152849..152995,155710..155811,156987..157068,
157392..157587,158401..158490,159714..159823,
160536..160613,161726..161797,164422..164482,
165754..165823,166414..166518,167248..167307,
167936..168012,172845..172934,173533..176791)

```

```
165754. .165823,166414. .166518,167248. .167307,
167936. .168012,172845. .172934,173533. .173727)
/ gene="XNPEP2"
/codon_start=1
/product="membrane-bound aminopeptidase p"
/protein_id="AAG28480.1"
/db_xref="GI:11066157"
/translation="MARAHMGCCPWLVLACAWHCTKPVDLGGQVNRCSNPPYLP
VTVNTMSLTALQOMQTNLSAYILPGTDHANEVYIGQDERRAMITGTSAGTA
VVMKKAATWDSYVQAEQMCNWEHKEVGTPIVWLLTEIPAGRGVDFPL
LSIDTWESYDALQSNQLVSYITNLVDLVMSERPPVNPPIYALQEAFTGSTWSE
KVGSRQMQKHQVPTAVLLSALEETAWLFNLRASDIPNPFYVYTLIDDSIRLF
ANKSRFSSTSYLNSCTGCMVCQIEDYQVSDRSIQAYSIGDVIWIGTSYTWGLY
EMIPEKLVITDYSYVMTKAVNKSQOALLKASHRDVAVIYLVLEKNVPGTV
DEFSGAEIVDFEREEQSPSPSPETISASGNNALAHYSKELNKLSSDENYLLD
SGGQVADGTTDITRVHMGTSFAQKQAYTRVLIGNIDLSRLIPPAATSGRMVFAFAR
RALMDALVYEAKTYPGSYLTFFVSVFVYDRLNIDVLSPEHLQVLRNYQVIREK
VGPELQRQLLEPEWMLQOHTPELAARAPDTASVSLVSTLALGWSV"
misc_feature
160606. .160608
/ gene="XNPEP2"
/ note="Putative proton shuttle; unclassified site"
161778. .161780
/ gene="XNPEP2"
/ note="divalent metal ligand; metal-binding site"
164435. .164437
/ gene="XNPEP2"
/ note="divalent metal ligand; metal-binding site"
166473. .166475
/ gene="XNPEP2"
/ note="divalent metal ligand; metal-binding site"
167977. .167979
/ gene="XNPEP2"
/ note="divalent metal ligand; metal-binding site"
175728. .176791
/ gene="XNPEP2"
/ note="divalent metal ligand; metal-binding site"
3'UTR
ORIGIN
Query Match 100.0%; Score 467; DB 9; Length 206618;
Best Local Similarity 100.0%; Pred. No. 6,8e-113;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTGCTATCCACCTCCCGCCCAAGCCCTTTACCTATGCTGCTGCTGCTG 60
Db 51518 CAATGTTGCTATCCACCTCCCGCCCAAGCCCTTTACCTATGCTGCTGCTGCTG 51459
QY 61 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db 51458 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
QY 121 GTCTCTTCTGCCACTGAGCCCCCATCAGGAGTTGGGCCCTTCTTCCCTCCCTTCTG 180
Db 51398 GTCTCTTCTGCCACTGAGCCCCCATCAGGAGTTGGGCCCTTCTTCCCTCCCTTCTG 51339
QY 181 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
Db 51338 TGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGAGATAGGGAACAGAGAGGGTGTGGTG 51279
Db 51278 GAGAAAGTGGGGGATGGCTAAGAAAGCTGAGATAGGGAACAGAGAGGGTGTGGTG 300
QY 301 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTATTAATTAACAA 360
Db 51218 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTATTAATTAACAA 51159
QY 361 AGATATTAAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTCAATGTTGTA 420
Db 51158 AGATATTAAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTCAATGTTGTA 51099
QY 421 AGATTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467
Db 51098 AGATTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 51052
RESULT 8
AX877684 2243 bp DNA linear PAT 17-DEC-2003
LOCUS Sequence 12589 from Patent EPI074617.
DEFINITION AX877684
ACCESSION AX877684
VERSION AX877684.1 GI:40032420
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primers for synthesising full-length cDNA and their use
Patent: Ep 1074617-A 12589 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
source
1. .2243
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
6. .374
/ note="unnamed protein product"
/codon_start=1
/protein_id="CAE89834.1"
/db_xref="GI:40032421"
/translation="MSATWCSPEGQMGQPGREVGNSSAAGPASPPIRNPLSEAGL
KGPPSAHLRLCLLHLVCFSGGLTSLQSPRTCCSHQWALFSPACFPQWRAPGCSL
DDSRSLTRIPRVHLPGPSLD"
ORIGIN
Query Match 97.2%; Score 454; DB 6; Length 2243;
Best Local Similarity 99.4%; Pred. No. 1.6e-109;
Matches 467; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
QY 1 CAATGTTGCTATCCACCTCCCGCCCAAGCCCTTTACCTATGCTGCTGCTGCTGCTG 57
Db 1774 CAATGTTGCTATCCACCTCCCGCCCAAGCCCTTTACCTATGCTGCTGCTGCTGCTG 1833
QY 58 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 117
Db 1834 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1893
QY 118 AAAGTCTCTTCTGCCACTGAGCCCCCATCAGGAGTTGGGCCCTTCTTCCCTCCCTTCT 177
Db 1894 AAAGTCTCTTCTGCCACTGAGCCCCCATCAGGAGTTGGGCCCTTCTTCCCTCCCTTCT 1953
QY 178 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237
Db 1954 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2013
QY 238 GGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGAGATAGGGAACAGAGAGGGTGTAGTGG 297
Db 2014 GGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGAGATAGGGAACAGAGAGGGTGTAGTGG 2073
QY 298 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTATTAATTTATA 357
Db 2074 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTATTAATTTATA 2133
QY 358 CAAAGATATTAAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTCAATGTTG 417
Db 2134 CAAAGATATTAAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTCAATGTTG 2193
QY 418 TAAAGATTTGTTCTGTTGTTAAATGTTCTTTATTAATAAAGAGTGTAAAGCTG 467
Db 2194 TAAAGATTTGTTCTGTTGTTAAATGTTCTTTATTAATAAAGAGTGTAAAGCTG 2243
```

```

RESULT 9
BD156791
LOCUS          2243 bp      DNA      linear      PAT 17-JAN-2003
DEFINITION    Primer for synthesizing full-length cDNA and use thereof.
ACCESSION    BD156791
VERSION      BD156791.1 GI:27862549
KEYWORDS     JP 2002191363-A/11634.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 2243)
AUTHORS     Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
              Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE       Primer for synthesizing full-length cDNA and use thereof
JOURNAL     HELIX RESEARCH INSTITUTE
            Patent: JP 2002191363-A 11634 09-JUL-2002;
COMMENT     OS Homo sapiens (human)
            PN JP 2002191363-A/11634
            PD 09-JUL-2002
            PF 28-JUL-2000 JP 2000280990
            PI TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI SAITO,
            PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI KEIICHI NAGAI,TETSUJI OTSUKI
            PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
            PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
            Primer for synthesizing full-length cDNA and use thereof FH Key
            Location/Qualifiers
            FT CDS
            source      1..2243
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"

ORIGIN
Query Match          97.2%; Score 454; DB 6; Length 2243;
Best Local Similarity 99.4%; Pred. No. 1.6e-109;
Matches 467; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 CAATGTTTGGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTGCTAAAG---CTG 57
DB 1774 CAATGTTTGGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTGCTAAAGCTACTG 1833

QY 58 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 117
DB 1834 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 1893

QY 118 AAAGTCTTCTTGCCACTGACGCCCCCATCAGGGATTGGGCTTCTTTCCCTTCCCTTT 177
DB 1894 AAAGTCTTCTTGCCACTGACGCCCCCATCAGGGATTGGGCTTCTTTCCCTTCCCTTT 1953

QY 178 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 237
DB 1954 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 2013

QY 238 GGGGAGAAAGTGGGAGTGGCTTAAAGAGCTGGAGATAGGGAACAGAGAGGAGTGTGG 297
DB 2014 GGGGAGAAAGTGGGAGTGGCTTAAAGAGCTGGAGATAGGGAACAGAGAGGAGTGTGG 2073

QY 298 GTGGGCTAGGGGGCTGCTTATTAAAGTGGTGTGTTATGATTTCTTAACTAAATTATA 357
DB 2074 GTGGGCTAGGGGGCTGCTTATTAAAGTGGTGTGTTATGATTTCTTAACTAAATTATA 2133

QY 358 CAAAGATATTAGGCCCTGTTTAAAGAAATGTTCCCTTCCCTGTTCAATGTTTG 417
DB 2134 CAAAGATATTAGGCCCTGTTTAAAGAAATGTTCCCTTCCCTGTTCAATGTTTG 2193

QY 418 TAAAGATTGTTCTGTGTAATATGCTTTTATAATAACAGTTAAAGCTG 467

```

```

DB 2194 TAAAGATTGTTCTGTGTAATATGCTTTTATAATAACAGTTAAAGCTG 2243

RESULT 10
AK001855
LOCUS          2243 bp      mRNA      linear      PRI 01-AUG-2002
DEFINITION    Homo sapiens cDNA FLJ10993 fis, clone PLACE1002140.
ACCESSION    AK001855
VERSION      AK001855.1 GI:7023382
KEYWORDS     oligo capping; fis (full insert sequence).
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1
AUTHORS     Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
              Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
              Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K.,
              Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
              Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Oshima,A.
              NEDO human cDNA sequencing project
              Unpublished
              2 (bases 1 to 2243)
              Isogai,T. and Otsuki,T.
              Direct Submission
              Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
              Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
              NEDO human cDNA sequencing project supported by Ministry of
              International Trade and Industry of Japan; cDNA full insert
              sequencing; Research Association for Biotechnology; cDNA library
              construction, 5'- & 3'-end one pass sequencing and clone selection;
              Helix Research Institute (supported by Japan Key Technology Center
              etc.) and Department of Virology, Institute of Medical Science,
              University of Tokyo.
              Location/Qualifiers
              1..2243
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="PLACE1002140"
              /tissue_type="placenta"
              /clone_lib="PLACE1"
              /note="cloning vector: pME18SFL3"
              6..374
              /note="unnamed protein product"
              /codon_start=1
              /protein_id="BAA91944.1"
              /db_xref="GI:7023383"
              /translation="MSATWCSPEGGMGQGPREVGNSSAAGSPASPIENPCLSEAGL
              KGPPSAHPLRLCLLHLVCFSGGLTSLSPRTCCSHQWQLFSPACFPQWRAPGCSL
              DDSRLRIRPVLHPGFSLD"

CDS
Query Match          97.2%; Score 454; DB 9; Length 2243;
Best Local Similarity 99.4%; Pred. No. 1.6e-109;
Matches 467; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 CAATGTTTGGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTGCTAAAG---CTG 57
DB 1774 CAATGTTTGGCTATCCACCTCCCAAGCCCTTTTACCTATGCTGCTGCTAAAGCTACTG 1833

QY 58 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 117
DB 1834 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGGTGCCCGAG 1893

QY 118 AAAGTCTTCTTGCCACTGACGCCCCCATCAGGGATTGGGCTTCTTTCCCTTCCCTTT 177
DB 1894 AAAGTCTTCTTGCCACTGACGCCCCCATCAGGGATTGGGCTTCTTTCCCTTCCCTTT 1953

QY 178 CTGCTGCTGCTGCTCATCGGCTGCTGCGGCTGCGGAGCCAGCCCGCTGGGGAA 237

```

[illegible]

Db TTAAGAAATTTGTTCCCTTCCCTGTGTTCAATGTTTGAAGAATGTTCTGTGTAATAT 39

QY 441 GTCTTTATAATAAACAGTTAAAGCTG 467

Db 38 GTCTTTATAATAAACAGTTAAAGCTG 12

RESULT 12  
AX332852/c

LOCUS AX332852 458 bp DNA linear PAT 09-JAN-2002

DEFINITION Sequence 3361 from Patent WO0194629.

ACCESSION AX332852

VERSION AX332852.1 GI:18123486

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE

AUTHORS Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Young, P.E., Augustus M., Carter, K.C., Ebner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.

ABSTRACT Cancer gene determination and therapeutic screening using signature gene sets

JOURNAL Patent: WO 0194629-A 3361 13-DEC-2001; Avalon Pharmaceuticals (US)

FEATURES

source Location/Qualifiers

1..458

/organism="Homo sapiens"

/mol\_type="unassigned DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 89.8%; Score 419.2; DB 6; Length 458;

Best Local Similarity 97.3%; Pred. No. 2.6e-100;

Matches 435; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 22 CCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTTAAA 81

Db 458 CCCCAAGCCCTTTACCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTAAA 399

QY 82 GGCT-CATGCTGGAGTGGGAGTGGTGGTGGCCAGAAAGTCTCTTGCCACTGACGC 140

Db 398 GGCTCCATGCTGGAGTGGGAGTGGTGGTGGCCAGAAAGTCTCTTGCCACTGACGC 339

QY 141 CCCCATCAGGAGTTGGCCCTTCTTCCCTCCCTTCTTCTGCTGCTGCTGCTGCTGCTG 200

Db 338 CCCCATCAGGAGTTGGCCCTTCTTCCCTCCCTTCTTCTGCTGCTGCTGCTGCTGCTG 279

QY 201 TGCATGACCTCAGCCAGCCAGCCCGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 260

Db 278 TGCATGACCTCAGCCAGCCAGCCCGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGG 219

QY 261 AGAAAGCTGGGAGATAGGAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 320

Db 218 AGAAAGCTGGGAGATAGGAGAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 159

QY 321 TTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 380

Db 158 TTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 99

QY 381 TTAAAGAAATTTGTTCCCTTCCCTGTTTCAATGTTTGAAGATTTGTTCTGTGTAATAT 440

Db 98 TTAAAGAAATTTGTTCCCTTCCCTGTTTCAATGTTTGAAGATTTGTTCTGTGTAATAT 39

QY 441 GTCTTTATAATAAACAGTTAAAGCTG 467

Db 38 GTCTTTATAATAAACAGTTAAAGCTG 12

RESULT 13  
AX874329/c

LOCUS AX874329 403 bp DNA linear PAT 17-DEC-2003

```

DEFINITION   Sequence 9234 from Patent EP1074617.
ACCESSION    AX874329
VERSION      AX874329.1  GI:40029100
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS      Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
            Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE        Primers for synthesizing full-length cDNA and their use
JOURNAL      Patent: EP 1074617-A 9234 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES     Location/Qualifiers
             source          1..403
                        /organism="Homo sapiens"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:9606"
ORIGIN
Query Match      83.8%; Score 391.2; DB 6; Length 403;
Best Local Similarity 97.5%; Pred. No. 7.1e-93;
Matches 393; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 65 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGTCGTCGCCAGAAAGTCT 124
DB 403 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGANTGGTCGGTCCCGAGAAAGTTT 344
QY 125 CTTCTGCCACTGACGCCGCCCATCAGGAGTGGGCTCTTTTCCCCCTTCCTTCTGTGTC 184
DB 343 TTTTTCGCCACTGACGCCGCCCATCAGGAGTGGGCTTTTCCCCCTTCCTTCTGTGTC 284
QY 185 TCCTGCTCATCGCCCTGCCATGACCTGCAGCCAGCCAGCCGCCGTCGGGAGGGGAGA 244
DB 283 TCCTGCTCATCGCCCTGCCATGACCTGCAGCCAGCCAGCCGCCGTCGGGAGGGGAGA 224
QY 245 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGTGGGCT 304
DB 223 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGTGGGCT 164
QY 305 AGGGGGCTGCTTAAAGTGGTGTATGATTTTATGATTTTATCTAATTTATACAAAGAT 364
DB 163 AGGGGGCTGCTTAAAGTGGTGTATGATTTTATGATTTTATCTAATTTATACAAAGAT 104
QY 365 ATTAAGCCCTGTTCAATAGAAATGTTCCCTTCCCTGTCATGTTGTAAGAT 424
DB 103 ATTAAGCCCTGTTCAATAGAAATGTTCCCTTCCCTGTCATGTTGTAAGAT 44
QY 425 TGTTCGTGTAATATGCTTTTATATAAAGTAAAGCTG 467
DB 43 TGTTCGTGTAATATGCTTTTATATAAAGTAAAGCTG 1

RESULT 14
LOCUS        BD154391/c
DEFINITION   Primer for synthesizing full-length cDNA and use thereof.
ACCESSION    BD154391
VERSION      BD154391.1  GI:27860149
KEYWORDS     JP 2002191363-A/9234
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS      Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
            Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE        Primers for synthesizing full-length cDNA and use thereof
JOURNAL      Patent: JP 2002191363-A 9234 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT      OS Homo sapiens (human)
            PN JP 2002191363-A/9234

```

```

PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
PI SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source          1..403
FT Location/Qualifiers
                        /organism="Homo sapiens (human)"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
ORIGIN
Query Match      83.8%; Score 391.2; DB 6; Length 403;
Best Local Similarity 97.5%; Pred. No. 7.1e-93;
Matches 393; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 65 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGTCGTCGCCAGAAAGTCT 124
DB 403 TGCCTGCTGCTTAAAGGCTCATGCTGGAGTGGGANTGGTCGGTCCCGAGAAAGTTT 344
QY 125 CTTCTGCCACTGACGCCGCCCATCAGGAGTGGGCTCTTTTCCCCCTTCCTTCTGTGTC 184
DB 343 TTTTTCGCCACTGACGCCGCCCATCAGGAGTGGGCTTTTCCCCCTTCCTTCTGTGTC 284
QY 185 TCCTGCTCATCGCCCTGCCATGACCTGCAGCCAGCCAGCCGCCGTCGGGAGGGGAGA 244
DB 283 TCCTGCTCATCGCCCTGCCATGACCTGCAGCCAGCCAGCCGCCGTCGGGAGGGGAGA 224
QY 245 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGTGGGCT 304
DB 223 AAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGTGGGCT 164
QY 305 AGGGGGCTGCTTAAAGTGGTGTATGATTTTATGATTTTATCTAATTTATACAAAGAT 364
DB 163 AGGGGGCTGCTTAAAGTGGTGTATGATTTTATGATTTTATCTAATTTATACAAAGAT 104
QY 365 ATTAAGCCCTGTTCAATAGAAATGTTCCCTTCCCTGTCATGTTGTAAGAT 424
DB 103 ATTAAGCCCTGTTCAATAGAAATGTTCCCTTCCCTGTCATGTTGTAAGAT 44
QY 425 TGTTCGTGTAATATGCTTTTATATAAAGTAAAGCTG 467
DB 43 TGTTCGTGTAATATGCTTTTATATAAAGTAAAGCTG 1

RESULT 15
LOCUS        AC127934
DEFINITION   Rattus norvegicus clone CH230-32806, WORKING DRAFT SEQUENCE.
ACCESSION    AC127934
VERSION      AC127934.3  GI:25075008
KEYWORDS     HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE       Rattus norvegicus (Norway rat)
ORGANISM     Rattus norvegicus
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
AUTHORS      1 (bases 1 to 171310)
            Muzny, D., Marie, Metzker, M., Lee, S., Adams, C., Alder, J.,
            Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
            Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
            Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
            Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
            Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

```

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, B., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyile, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabris, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, B., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. I., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulsegged, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelimeh, O., Okwuonu, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Sisson, I., Sitter, C. D., Smajd, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valae, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission  
Unpublished  
2 (bases 1 to 171310)  
Worley, K. C.  
Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 171310)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23915370.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: KAAV  
Center clone name: CH230-32806  
----- Summary Statistics  
Assembly program: Phrap, version 0.990329  
Consensus quality: 163795 bases at least Q40  
Consensus quality: 165400 bases at least Q30  
Consensus quality: 166622 bases at least Q20  
Estimated insert size: 169296; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 171310: contig of 171310 bp in length.  
\* Location/Qualifiers  
1..171310  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-32806"  
1..1280  
/notes="wgs end extension  
clone\_end:Sp6"  
1567..3587  
/notes="wgs end extension  
clone\_end:Sp6"  
5295..6132  
/notes="clone boundary  
clone\_end:Sp6"  
site:  
end sequence:BZ164045"  
168396..169235  
/notes="clone boundary  
clone\_end:T7"  
site:  
end sequence:BZ164044"  
169684..171310  
/notes="wgs end extension  
clone\_end:T7"  
Query Match 32.5%; Score 151.8; DB 2; Length 171310;  
Best Local Similarity 63.3%; Pred. No. 3.3e-29;  
Matches 307; Conservative 0; Mismatches 157; Indels 21; Gaps 4;  
QY 4 TGTTCCTTATCCACTCCCAAGCCCTTTACCTATGCTGCTGTAAGCTGCTG 63  
Db 116761 TGATTGCTATTTCACCCGTCCTCAAAATCCCATCTGCTGCTGCTGCTG 116820  
QY 64 CTGCTGCTGCTGCTTAAAGGCTCATCTTGGAGTGGGACTGCTGGTCCCAAGAAATC 123  
Db 116821 CTGCTGCTGCTTCAAGGCCCACTGGGAGCTGGGCCAGTCATAGATCCTCCTT 116880  
QY 124 TCTTTCGCCATGACGCCGCCCATCAGGAGTTGGGCTTCTTCCCTTC----- 173  
Db 116881 GGGCTTGAAGTGGGGCTGCTGGTGGTCCCTTCTACTGCTGGTACCCCTGCCAGGACCA 116940  
QY 174 -----CTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 227  
Db 116941 GGCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 117000  
QY 228 CCGTGGGGAAGGGGA--GAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGA 285



```

Db      117001 CCACAGGGAAGGGAGGGAAGGAGTGTGGCTGAGAAAGAGAGATAGAAACAGAA 117060
QY      286 AGAGGGTAGTGGGTGGGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATTCCTTA 345
Db      117061 GAGGGGAGTGAATGGAACCCAGTGGGTGTCTTATTTAAAGTGGTGTATGATTCCTTA 117120
QY      346 TACTAATTTATACAAAGATATTAGGCCCTGTTTCATT-AAGAAATGTTCCCTTCCCTG 404
Db      117121 TACTAATTTATAGAGATATTAGGCCCTGCTGAGTTAAAGAACTGCTCATCCCGTG 117180
QY      405 TGTTCA--ATGTTTGTAAAGATTGTTCTGTGTAATAATGCTTTATATATAACAGTTTAAA 462
Db      117181 TGTTCACTATGTTTGTAAAAAATGTTCCATGTAAATATGTTCTTATATAAGAGTTTAAA 117240
QY      463 AGCTG 467
Db      117241 AGCTG 117245

RESULT 16
AC136814
LOCUS   Rattus norvegicus clone CH230-179117, *** SEQUENCING IN PROGRESS
DEFINITION
AC136814
VERSION AC136814.2 GI:25188346
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE  Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 208810)
Muzny,D,Marie., Metzker,M,Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Altsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blaukenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,K., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindarne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidas,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,

```

```

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 208810)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (08-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 208810)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:24796687.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KCOW
Center clone name: CH230-179117
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 192829 bases at least Q40
Consensus quality: 195654 bases at least Q30
Consensus quality: 197508 bases at least Q20
Estimated insert size: 201423; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 208810: contig of 208810 bp in length.
Location/Qualifiers
1..208810
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-179117"
complement(588..1486)
/note="clone_boundary
misc_feature
source
FEATURES

```



```

clone_end:sp6
site:EcoRI
end sequence:RH303035"
161506...163638
/notes="wgs contig"
176821...176860
/notes="clone boundary
clone_end:T7
site:EcoRI
end sequence:RH303034"
178559...181095
/notes="wgs end extension
clone_end:T7"

ORIGIN
Query Match      32.5%; Score 151.8; DB 2; Length 208810;
Best Local Similarity 63.3%; Pred. No. 3.3e-29;
Matches 307; Conservative 0; Mismatches 157; Indels 21; Gaps 4;

QY 4 TGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAGCTGCTGCTG 63
   |||||
Db 34766 TGAATGCTATTACCGCTCCCAAAATCCTTCCCATCTGCTGCTAATGCTGCTG 34825
   |||||

QY 64 CTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGCTGCTGCCAGAAAGTC 123
   |||||
Db 34826 CTGCTGCGCTTCAAGGCCACCCCTGGGAGCTGGGCGAGTCATAGAGTCTCACTCCTT 34885
   |||||

QY 124 TCYTTCGCACTGACGCCCCCATCAGGAGTGGGCTTCTTCCCTTC----- 173
   |||||
Db 34886 GGGCTTGAATGGGGGCTTCTGGTGCTTCTTCTACTGCTGGTACCCCTGCCAGGGACCA 34945
   |||||

QY 174 -----CTTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 227
   |||||
Db 34946 GGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 35005
   |||||

QY 228 CCGTGGGGAAGGGA--GAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGAACAGA 285
   |||||
Db 35006 CCACAGGGAAGGGAGGGAAGGAGGATGTTGGCTGAGAAAGAGAGAGATAGAAACAGAA 35065
   |||||

QY 286 AGAGGGTAGTGGGTGGGCTAGGGGGGCTGCTTATTTAAAGTGGTCTTTATGATCTTAA 345
   |||||
Db 35066 GAGGGGAGTGAATGGAACCAAGTGGGCTGTCTTATTTAAAGTGGTGTGTATGATCTTAA 35125
   |||||

QY 346 TACTAATTTATACAAAGATATTAGGCCCTGTTTCATT-AGAAATTTCTCCCTCCCGTG 404
   |||||
Db 35126 TACTAATTTATAGAGATATTAGGCCCTCTGAGTTAAAGAACTGTCTCTATCCCGTG 35185
   |||||

QY 405 TGTTCA--ATGTTTGAAGAATGTTCTGTGTAATAATGCTTTTAAATAAACAGTTAAA 462
   |||||
Db 35186 TGTTCACTATGTTTGTAAAAAATGTTCCATGTAATAATGCTTTTAAATAAAGAGTTAAA 35245
   |||||

QY 463 AGCTG 467
   |||||
Db 35246 AGCTG 35250
   |||||

```

```

RESULT 17
BC020015
LOCUS
DEFINITION Mus musculus apelin, mRNA (cDNA clone MGC:27819 IMAGE:3483588),
complete cds.
ACCESSION BC020015
VERSION BC020015.1 GI:18043669
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3143)
REFERENCE
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Hen, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Atschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

```

```

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Udín, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Rahe, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
22388257
12477932
2 (bases 1 to 3143)
Strausberg, R.
Direct Submission
Submitted (19-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapsb-f@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 35 Row: f Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7304894.
FEATURES
Location/Qualifiers
1..3143
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="MGC:27819 IMAGE:3483588"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI CGAP Mam5"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..3143
/gene="Apelin"
/notes="synonym: Apelin"
/db_xref="LocusID:30878"
/db_xref="MGI:1353624"
329..562
/codon_start=1
/product="apelin"
/protein_id="AAH20015.1"
/db_xref="GI:18043670"
/db_xref="LocusID:30878"
/translation="MNLRLCVQALLLLSLTAVCGVPLMPPDGTGLESGMRYLVK
PRTSRVGPAGWQGRKFRQRRLSHKGMPFP"
ORIGIN

```



regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-351G20 is from the RPCI-23 Mouse PAC library  
constructed by the group of Pieter de Jong.  
For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

## FEATURES

Location/Qualifiers

```
1..189982
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="X"
    /clone="RP23-351G20"
    /clone_lib="RPCI-23"
```

## ORIGIN

```
Query Match      27.8%; Score 129.8; DB 10; Length 189982;
Best Local Similarity 69.0%; Pred. No. 2.3e-23;
Matches 245; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

QY 121 GTCTCTCTGCACTGACGCCCATCAGGATTTGGCCCTTCTTCCCTTCTCTCTTCT 179
Db 1696 GCCTCTCTCTGCTGTGTACCTTACAGGACCGGCGCTTCTCTCTTGTGCTTCT 1637

QY 180 GTGTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 239
Db 1636 GCACCTCTGCTTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1577

QY 240 GGAGAAAGTGGGATGGCTAAGAAAGTGGGATAGGACAGAGAGAGTGTGGGT 299
Db 1576 T-----TGCTGAGAAAGAGGAGATAGA-AGCAAGAGGGGAGTAAAC 1532

QY 300 GGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGTTCTTATCTTACTAATTTATACA 359
Db 1531 GGGCCAGTGGCTGTCTTATTTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1472

QY 360 AAGATATTAAGCCCTGCTGCTTAAAGAAATGTTTCCCTTCCCTTCCCTTCCCTTCCCT 412
Db 1471 GAGATATTAAGCCCTTGAAGTGAAGAACTGTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 1412

QY 413 GTTTGTAAGATTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467
Db 1411 GTTTGTAAGATTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1357
```

```
RESULT 20
AC055817 AC055817 203805 bp DNA linear ROD 20-JAN-2003
LOCUS Mus musculus, clone RP23-15417, complete sequence.
DEFINITION AC055817
ACCESSION AC055817
VERSION AC055817.12 GI:27804302
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 203805)
```

Authors Birren, B., Nusbaum, C. and Lander, E.

Title Mus musculus, clone RP23-15417

Journal Unpublished

Reference 2 (bases 1 to 203805)

Authors Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, N., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McGowan, P., McGurk, A., McKernan, K., McPhee, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 203805)

## REFERENCE

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gort, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-DEC-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 203805)

## REFERENCE

## AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cooke, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (20-JAN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jan 20, 2003 this sequence version replaced gi:26225143.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

RESULT 21	
HUMZD38F03	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL

COMMENT

FLI CDNA.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 598)  
Woessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,  
Marth,G., Bowers,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,  
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,  
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,  
Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behnmer,K., Hillier,L.,  
Wilson,R. and Waterston,R.  
Full Clone Sequencing of the Longest Available Member from Each  
Unigene Cluster  
Unpublished  
2 (bases 1 to 598)  
Waterston,R.  
Direct Submission  
Submitted (24-AUG-1998) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
SUBMITTED BY:  
Genome Sequencing Center  
Department of Genetics  
Washington University  
St. Louis MO 63108, USA  
http://genome.wustl.edu/gsc  
mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES  
source

ORIGIN

Query Match  
Best Local Similarity 100.0%; Score 127; DB 9; Length 598;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 TCTTATACATAATTTATACAAAGATTAAAGCCCTGTTTCATTAAAGAAATTTCCCTTC 400  
Db 1 TCTTATACATAATTTATACAAAGATTAAAGCCCTGTTTCATTAAAGAAATTTCCCTTC 60

Qy 401 CCTGTGTTCAATGTTGTAAGATGTTCTGTGTAATATGTTCTTTATAAACAAGTTA 460  
Db 61 CCTGTGTTCAATGTTGTAAGATGTTCTGTGTAATATGTTCTTTATAAACAAGTTA 120

Qy 461 AAAGCTG 467  
Db 121 AAAGCTG 127

RESULT 22  
AX211242  
LOCUS  
DEFINITION  
ACCESSION  
VERSION

AX211242  
Sequence 884 from Patent WO0157058.  
AX211242  
AX211242.1 GI:15425502

KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL

FEATURES  
source

ORIGIN

Query Match  
Best Local Similarity 15.9%; Score 74.2; DB 6; Length 255;  
Matches 168; Conservative 8; Mismatches 66; Indels 9; Gaps 5;

Qy 199 CCTGCCATGACCTGCAGCCCAAGCCCGCCGTCGGGGAAGGGGA--GAAAGTGGGGATG 256  
Db 5 CTTCCTTATCAGCCTGCATGCCCAATCCACAGGGAACSSGAGGGAAGGAGATGTTG 64

Qy 257 GCTAAGAAAGCTGGGAGATAGGGAAACAGAGAGAGGTAGTGGCTAGGGGGGCTGCC 316  
Db 65 GCTGASAAASMSGAGAGATASAVACAGAGAGGGGGAGTGAATGGACCCGCTGTC 124

Qy 317 TTATTT-AAAGTGGTCTTATGATCTTATACATA-ATTATACAAAGATATTAAAGCCC 374  
Db 125 TTATTTCAAGTGGTGTGTATGATTTCTTATCTATATAGATATTAAAGCCC 184

Qy 375 TGTTCATT-AAGAAATTTTCCTTCCTG---TGTTCAAATGTTTAAAGATTTGTC 429  
Db 185 TCTGAGTTAAGAAACTSYCTCATCCGGTCTGTTCACTCATGTTTGTAAAAATTTGTC 244

Qy 430 TGTGTAATAT 440  
Db 245 CATGCTAACAT 255

RESULT 23  
AX344562  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AX344562  
Sequence 13 from Patent WO0200932.  
AX344562  
AX344562.1 GI:18492448  
synthetic construct  
synthetic construct  
artificial sequences.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Epigenomics AG (DE)  
Location/Qualifiers  
1. 73778  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length

of seq 2: 3.673778 <223>--split as follows--seq 14  
 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16  
 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001  
 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001  
 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001  
 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001  
 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001  
 3.673.778"

## ORIGIN

Query Match 9.9%; Score 46.4; DB 6; Length 73778;  
 Best Local Similarity 50.0%; Pred. No. 0.31;  
 Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 230 GTGGGAAGGGAGAAAGTGGGGGATGGCTACAAAGCTGGGAGATAGGGAACAGAGAG 289  
 DB 42743 GAGTGGAACTTTGGAGGGGAGGGGTTTCTAAATTTATTTGGAAGATTTGTGATTGAAAAGG 42802

QY 290 GGTAGTGGCTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTTCATTAATCTTATCT 349  
 DB 42803 ATTGAGGAATGGGTAAAGAGATCGTTTTTTTATAGTATTCATTTATATTTATTTT 42862

QY 350 AATTATACAAAGATTAAGGCCCTGTTCATTAAGAAATGTTCCTTCCTCCCTGTGTTTC 409  
 DB 42863 TTTTGTGTTAGTTTATTTAGGAGAGATTAAGAGAGAAATGTTTGTGTTT 42922

QY 410 AATGTTTGAAGATTTCTGTGTAATATGCTCTTTATATAAACAGTTAA 461  
 DB 42923 TATGTCGGATGTTTTTTTTTGTGGAGTTTTTTTATTTGGAGATGGAGTATTTA 42974

## RESULT 24

AX344561  
 LOCUS AX344561 349980 bp DNA linear PAT 01-FEB-2002  
 DEFINITION Sequence 12 from Patent WO200932.  
 ACCESSION AX344561  
 VERSION AX344561.1 GI:18492447  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.

REFERENCE 1  
 Olek,A., Piepenbrock,C. and Berlin,K.  
 TITLE Diagnosis of Known genetic parameters within the mbc  
 JOURNAL Patent: WO 020932-A 12 03-JAN-2002;  
 Epigenomics AG (DE)

FEATURES  
 Location/Qualifiers  
 1..349980  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>--split as follows--seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>--Original length of seq 2: 3.673778 <223>--split as follows--seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001 3.673.778"

FEATURES  
source

QY 230 GTGGGAAGGGAGAAAGTGGGGGATGGCTACAAAGCTGGGAGATAGGGAACAGAGAG 289  
 DB 342743 GAGTGGAACTTTGGAGGGGAGGGGTTTCTAAATTTATTTGGAAGATTTGTGATTGAAAAGG 342802

QY 290 GGTAGTGGCTGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTTCATTAATCTTATCT 349  
 DB 342803 ATTGAGGAATGGGTAAAGAGATCGTTTTTTTATAGTATTCATTTATATTTATTTT 342862

QY 350 AATTATACAAAGATTAAGGCCCTGTTCATTAAGAAATGTTCCTTCCTCCCTGTGTTTC 409  
 DB 342863 TTTTGTGTTAGTTTATTTAGGAGAGATTAAGAGAGAAATGTTTGTGTTT 342922

QY 410 AATGTTTGAAGATTTCTGTGTAATATGCTCTTTATATAAACAGTTAA 461  
 DB 342923 TATGTCGGATGTTTTTTTTTGTGGAGTTTTTTTATTTGGAGATGGAGTATTTA 342974

## RESULT 25

AX344561  
 LOCUS AX344561 134971 bp DNA linear HTG 31-OCT-2002  
 DEFINITION Oryza sativa (japonica cultivar-group) chromosome 11 clone OSUNBa0059H21, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 5 ordered pieces.  
 ACCESSION AX344561  
 VERSION AC116367.9 GI:24431629  
 KEYWORDS HTG; HTGS PHASE2.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 134971)  
 Buell,C.R., Yuan,Q., Quyang,S., Liu,J., Gansberger,K., Jones,K.M., Overton II,L., Teitelin,T., Kim,M., Bera,J., Jin,S., Fadrosch,D.W., Tallon,L., Koo,H., Zismann,V., Hsiao,J., Blunt,S., Vanaken,S., Riedmuller,S.B., Utterbach,T., Feldblyum,T., Yang,Q., Haas,B., Suh,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and Fraser,C.  
 Oryza sativa ssp. japonica cv. Nipponbare OSUNBa0059H21 BAC genomic sequence

TITLE Oryza sativa ssp. japonica cv. Nipponbare OSUNBa0059H21 BAC genomic sequence

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 134971)  
 Buell,R.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (27-MAR-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE 3 (bases 1 to 134971)  
 Buell,R.  
 AUTHORS Direct Submission  
 JOURNAL Submitted (31-OCT-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

## COMMENT

On Oct 31, 2002 this sequence version replaced gi:23308041.  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and the accession number will be preserved.

1 29813: contig of 29813 bp in length  
 29814 29913: gap of unknown length  
 29914 53824: contig of 23911 bp in length  
 53825 53924: gap of unknown length  
 53925 71255: contig of 17331 bp in length  
 71256 71355: gap of unknown length  
 71356 115072: contig of 43717 bp in length  
 115073 115172: gap of unknown length  
 115173 134971: contig of 19799 bp in length.

## FEATURES

Location/Qualifiers

1..134971

/organism="Oryza sativa (japonica cultivar-group)"

## Query Match

Best Local Similarity 50.0%; Pred. No. 0.33;  
 Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

BOULE

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-402H13"
1..1286
note="wgs_end_extension"
clone_end:T7"
1816..2986
note="clone_boundary"
clone_end:T7"
site:
end_sequence:B2254305"
202874..204064
note="clone_boundary"
clone_end:Sp6
site:
end_sequence:B2254306"
207192..208575
note="wgs_end_extension"
clone_end:Sp6"
208626..211683
note="wgs_end_extension"
clone_end:Sp6"

ORIGIN
Query Match          9.6%; Score 45; DB 2; Length 2128897;
Best Local Similarity 58.6%; Pred. No. 0.76;
Matches 78; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      4  TGTGTGCTATCCACCTCCCCAGCCCTTACCTATGCTGCTGTAACGCTGCTGCTG 63
        |||||
Db      36657 TGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 36598

QY      64  CTGCTGCTGCTGCTTAAAGCTCATGCTTGAGTGGGCACTGCTGCTGCCAGAAAGTC 123
        |||||
Db      36597 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 36538

QY      124  TCTCTGCACTG 136
        |||||
Db      36537 GCTTCTGCTACTG 36525

RESULT 27
AC105482
LOCUS      272030 bp DNA linear HTG 09-NOV-2002
DEFINITION Rattus norvegicus clone CH230-20G17, WORKING DRAFT SEQUENCE, 13
unordered pieces.
ACCESSION AC105482.4 GI:24818814
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 272030)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,V., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flegg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgiev,G., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

```

```

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Loreshuwa,L., Loulseged,H., Lozard,R.J., Lu,X., Ma,J.,
Maheehwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Mirja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Flopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Fuazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartabeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmari,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G., and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 272030)
Worley,K.C.

Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 272030)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 9, 2002 this sequence version replaced gi:23267948.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/) Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNAU
Center clone name: CH230-20G17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 255866 bases at least Q40
Consensus quality: 259060 bases at least Q30
Consensus quality: 260838 bases at least Q20
Estimated insert size: 263005; sum-of-contigs estimation

```



Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 13 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

```

1
  5173: contig of 5173 bp in length
  5174: gap of unknown length
  5274: contig of 10539 bp in length
  15812: gap of unknown length
  15813: contig of 3681 bp in length
  15993: contig of 238804 bp in length
  19693: gap of unknown length
  19694: contig of 238804 bp in length
  258497: gap of unknown length
  258498: contig of 1210 bp in length
  258598: gap of unknown length
  25907: contig of 1022 bp in length
  260923: gap of unknown length
  261029: contig of 1158 bp in length
  262187: gap of unknown length
  262287: contig of 1158 bp in length
  262288: gap of unknown length
  263445: gap of unknown length
  263446: contig of 1427 bp in length
  263546: gap of unknown length
  264972: contig of 1002 bp in length
  265073: gap of unknown length
  266174: contig of 1298 bp in length
  266175: gap of unknown length
  267473: contig of 1298 bp in length
  267572: gap of unknown length
  267573: contig of 2255 bp in length
  269828: gap of unknown length
  269928: contig of 2103 bp in length.
FEATURES
  source
    1..272030
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clones="CH230-20617"
    1..1229
      /note="wgs end_extension"
      /clone_end:T7"
  misc_feature
    1969..2830
      /note="clone_boundary"
      /clone_end:T7"
      site:ECORI
      end_sequence:BH320006"
  misc_feature
    15913..17329
      /note="wgs_contig"
      237547..238337
      /note="clone_boundary"
      /clone_end:Sp6
      site:ECORI
      end_sequence:BH320007"

```

## ORIGIN

Query Match 9.6%; Score 45; DB 2; Length 272030;  
Best Local Similarity 58.6%; Pred. No. 0.77;  
Matches 78; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 4 TGTTCCTATCCACCTCCCGCCCTTTACTATGCTGCTGCTAAGCTGCTGCTG 63  
Db 64933 TGACAGCTGCTGCTCCCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 64992  
Qy 64 CTGCTGCTGCTGCTAAGGCTCATGCTGAGTGGGACTGCTGCTGCCAGAAAGTC 123  
Db 64993 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 65052

Qy 124 TCCTCTGCTGCTGCTG 136  
Db 65053 GCCTCTGCTGCTGCTG 65065

RESULT 28  
AC101159  
LOCUS  
DEFINITION Mus musculus clone RP23-173D8, LOW-PASS SEQUENCE SAMPLING.  
AC101159  
ACCESSION  
VERSION AC101159.1 GI:17059933  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 71781)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Mus musculus, clone RP23-173D8  
Unpublished  
2 (bases 1 to 71781)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukigalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Coepele,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,  
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,  
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
Lamazaras,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,  
MacCarthy,M., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunthang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L16056  
Center clone name: 173\_D\_8  
-----  
\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 737: contig of 737 bp in length  
\* 738 837: gap of 100 bp  
\* 838 1555: contig of 718 bp in length  
\* 1556 1655: gap of 100 bp

```
* 1656 2389: contig of 734 bp in length
* 2390 2489: gap of 100 bp
* 2490 3228: contig of 739 bp in length
* 3229 3328: gap of 100 bp
* 3329 4057: contig of 739 bp in length
* 4058 4157: gap of 100 bp
* 4158 4887: contig of 720 bp in length
* 4888 4987: gap of 100 bp
* 4988 5715: contig of 728 bp in length
* 5716 5815: gap of 100 bp
* 5816 6559: contig of 744 bp in length
* 6560 7403: contig of 744 bp in length
* 7404 7503: gap of 100 bp
* 7504 8249: contig of 746 bp in length
* 8250 8349: gap of 100 bp
* 8350 9071: contig of 722 bp in length
* 9072 9171: gap of 100 bp
* 9172 9906: contig of 735 bp in length
* 9907 10006: gap of 100 bp
* 10007 10749: contig of 743 bp in length
* 10750 10849: gap of 100 bp
* 10850 11600: contig of 751 bp in length
* 11601 11700: gap of 100 bp
* 11701 12457: contig of 757 bp in length
* 12458 12557: gap of 100 bp
* 12558 13307: contig of 750 bp in length
* 13308 13407: gap of 100 bp
* 13409 14123: contig of 722 bp in length
* 14130 14229: gap of 100 bp
* 14230 14935: contig of 706 bp in length
* 14936 15035: gap of 100 bp
* 15036 15768: contig of 733 bp in length
* 15769 15868: gap of 100 bp
* 15869 16604: contig of 736 bp in length
* 16605 16704: gap of 100 bp
* 16705 17439: contig of 735 bp in length
* 17440 17539: gap of 100 bp
* 17540 18275: contig of 736 bp in length
* 18276 18375: gap of 100 bp
* 18376 19100: contig of 725 bp in length
* 19101 19200: gap of 100 bp
* 19201 19938: contig of 738 bp in length
* 19939 20038: gap of 100 bp
* 20039 20788: contig of 750 bp in length
* 20789 20888: gap of 100 bp
* 20889 21630: contig of 742 bp in length
* 21631 21730: gap of 100 bp
* 21731 22457: contig of 727 bp in length
* 22458 22557: gap of 100 bp
* 22558 23308: contig of 751 bp in length
* 23309 24146: gap of 100 bp
* 24147 24246: gap of 100 bp
* 24247 24971: contig of 725 bp in length
* 24972 25071: gap of 100 bp
* 25072 25816: contig of 745 bp in length
* 25817 25916: gap of 100 bp
* 25917 26659: contig of 743 bp in length
* 26660 26759: gap of 100 bp
* 26760 27504: contig of 745 bp in length
* 27505 27604: gap of 100 bp
* 27605 28376: contig of 772 bp in length
* 28377 28476: gap of 100 bp
* 28477 29212: contig of 736 bp in length
* 29213 30041: contig of 729 bp in length
* 30042 30141: gap of 100 bp
* 30142 30873: contig of 732 bp in length
* 30874 30973: gap of 100 bp
* 30974 31708: contig of 735 bp in length
* 31709 31808: gap of 100 bp
* 31809 32507: contig of 699 bp in length
```

```
* 32508 32607: gap of 100 bp
* 32608 33340: contig of 733 bp in length
* 33441 34150: contig of 710 bp in length
* 34151 34251: gap of 100 bp
* 34251 34998: contig of 748 bp in length
* 34999 35098: gap of 100 bp
* 35099 35845: contig of 747 bp in length
* 35846 35945: gap of 100 bp
* 35946 36690: contig of 745 bp in length
* 36691 37531: contig of 741 bp in length
* 37532 37632: gap of 100 bp
* 37632 38368: contig of 737 bp in length
* 38369 38468: gap of 100 bp
* 38469 39124: contig of 656 bp in length
* 39125 39224: gap of 100 bp
* 39225 39964: contig of 740 bp in length
* 39965 40064: gap of 100 bp
* 40065 40798: contig of 734 bp in length
* 40799 40898: gap of 100 bp
* 40899 41647: contig of 749 bp in length
* 41648 41747: gap of 100 bp
* 41748 42492: contig of 745 bp in length
* 42493 42592: gap of 100 bp
* 42593 43313: contig of 721 bp in length
* 43314 43413: gap of 100 bp
* 43414 44157: contig of 744 bp in length
* 44158 44257: gap of 100 bp
* 44258 45006: contig of 749 bp in length
* 45007 45106: gap of 100 bp
* 45107 45864: contig of 758 bp in length
* 45865 45964: gap of 100 bp
* 45965 46713: contig of 749 bp in length
* 46714 46813: gap of 100 bp
* 46814 47557: contig of 744 bp in length
* 47558 47657: gap of 100 bp
* 47658 48384: contig of 727 bp in length
* 48385 48484: gap of 100 bp
* 48485 49221: contig of 737 bp in length
* 49222 49321: gap of 100 bp
* 49322 50043: contig of 722 bp in length
* 50044 50143: gap of 100 bp
* 50144 50880: contig of 737 bp in length
* 50881 50980: gap of 100 bp
* 50981 51702: contig of 722 bp in length
* 51703 51802: gap of 100 bp
* 51803 52511: contig of 709 bp in length
* 52512 52611: gap of 100 bp
* 52612 53350: contig of 739 bp in length
* 53351 53450: gap of 100 bp
* 53451 54182: contig of 732 bp in length
* 54183 54282: gap of 100 bp
* 54283 55030: contig of 748 bp in length
* 55031 55130: gap of 100 bp
* 55131 55830: contig of 700 bp in length
* 55831 55930: gap of 100 bp
* 55931 56688: contig of 758 bp in length
* 56689 56788: gap of 100 bp
* 56789 57551: contig of 763 bp in length
* 57552 57651: gap of 100 bp
```

Query Match 9.5%; Score 44.2; DB 2; Length 71781;

Best Local Similarity 61.9%; Pred. No. 1.2; Mismatches 0; Indels 0; Gaps 0;

Matches 70; Conservative 0;

QY 2 AATGTTTGCTATCCACCTCCCAAGGCCCTTTACCTATGCTGTGCTAAAGCTGCTGC 61

DB 16221 ATGGGGGGTTTCACTGTACAAAGGGGCCAGTTCTGCTGTGCTGCTGCTGC 16280

QY 62 TCTGCTGCTGCTGTAAAGCTCATGCTTGGAGTGGGACTGCTGGTGCC 114

DB 16281 TCTGCTGCTGCTGTAAAGCTAAAGTTGGAGTGGGATCTTCAGTTCC 16333



Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission  
Unpublished  
2 (bases 1 to 242655)  
Worley, K.C.  
Direct Submission  
Submitted (22-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 242655)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:22855692.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GIXA  
Center clone name: CH230-86H10  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 230159 bases at least Q40  
Consensus quality: 232877 bases at least Q30  
Consensus quality: 234876 bases at least Q20  
Estimated insert size: 245378; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 239139: contig of 239139 bp in length  
\* 239140 239239: gap of unknown length  
\* 239240 240581: contig of 1342 bp in length  
\* 240582 240681: gap of unknown length  
\* 240682 242655: contig of 1974 bp in length.  
Location/Qualifiers  
1..242655  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-86H10"  
1..1966  
/note="wgs\_contig"  
2365..5223  
/note="wgs contig"

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

[http://www.sanger.ac.uk/projects/c\\_elegans/wormpep](http://www.sanger.ac.uk/projects/c_elegans/wormpep) RP23-173D8 is from the RPCI-23 Mouse PAC library<sup>1</sup> constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6. Location/Qualifiers

FEATURES

Location/Qualifiers

Accession	Genebank	Accession	Genebank	Accession	Genebank
QY	11	CTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTGCTGCTGC	70		
Db	336	CTGTCGGGTCCTTGTTGTCGAGTCCATATCTCTCTGTCCTGCTGCTGCTGCTGCGCGGC	277		

QY 71 TGCTGCTTAAGGCTCATGCTTGGAGTGGGAGCTGGTCCGCCAAGAAAGTCTTCTG 130  
 Db |||||  
 276 GGCTGCTCAGCTCAGATGCCAGATGGTCCGATGGTCCATATCAGCTGGTCTTGTG 217  
 QY 131 CCATCGAGCGCCCAATCAGGAGTGGGCTTCTTCCGCCCTTCTCTTCTGTCCTCCGTC 190  
 Db |||||  
 216 TCCTGGTCTTCTGCTCGGGTCTTGTGTCGGGTCCATATCCTCTGCTGCTGCTG 157  
 QY 191 CTATCGGCTGCCAGACCTGCGAGCCAGCCAG 225  
 Db |||||  
 156 TCTGCGCGCGGCTGATGACCTCCAGGTCCAG 122

RESULT 34  
 AC133581/C  
 LOCUS  
 DEFINITION Mus musculus chromosome 1 clone RP24-395P3 map 1, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 6 unordered pieces.  
 AC133581  
 VERSION AC133581.4 GI:39979530  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 155665)  
 Birren,B., Nusbaum,C. and Lander,E.  
 Mus musculus chromosome 1, clone RP24-395P3  
 Unpublished  
 2 (bases 1 to 155665)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Gardyna,S., Gord,S., Graham,L., FitzGerald,M., Gage,D., Galagan,J.,  
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (14-SEP-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 155665)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
 Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeRellano,K.,  
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,  
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,  
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
 Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,  
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (17-DEC-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Dec 17, 2003 this sequence version replaced gi:28933642.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L27198  
 Center clone name: 395\_P\_3  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 6 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 12796: contig of 12796 bp in length  
 \* 12797 12896: gap of 100 bp  
 \* 12897 59863: contig of 46967 bp in length  
 \* 59864 59863: gap of 100 bp  
 \* 59864 93811: contig of 33848 bp in length  
 \* 93811 93811: gap of 100 bp  
 \* 93812 93911: gap of 100 bp  
 \* 93912 14271: contig of 20360 bp in length  
 \* 14272 14371: gap of 100 bp  
 \* 14372 130845: contig of 16474 bp in length  
 \* 130846 130946: gap of 100 bp  
 \* 130946 155665: contig of 24720 bp in length.  
 FEATURES  
 Location/Qualifiers  
 1. 155665  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /map="1"  
 /clone="RP24-395P3"  
 /clone\_lib="RPCI-24 Male Mouse BAC"  
 ORIGIN  
 Query Match 9.4%; Score 43.8; DB 2; Length 155665;  
 Best Local Similarity 76.1%; Pred. No. 1.6;  
 Matches 54; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
 QY 7 TTGCTATTCACCTCCGCCAAGCCCTTTACCTATGCTGCTGCTGCTGCTGCTG 66  
 Db 154263 TTCCAGTCCGCCCTCCCAACCGCGCAGCCTCTGCTGCTGCTGCTGCTG 154204  
 QY 67 CTGCTGCTGCT 77  
 Db 154203 CTGCTGCTGCT 154193  
 RESULT 35  
 HSDJ989D7  
 LOCUS  
 DEFINITION Human DNA sequence from clone RPS-989D7 on chromosome 20. Contains  
 the 3' end of a novel gene, STGs and GSSs, complete sequence.  
 ACCESSION AL121911  
 VERSION AL121911.15 GI:9864671  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 119790)  
 Griffiths,C.

TITLE  
JOURNAL

Direct Submission

Submitted (02-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk

## COMMENT

On Aug 22, 2000 this sequence version replaced gi:9714378.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information  
on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone  
RP5-989D7. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.

The true left end of clone RP5-989D7 is at 1 in this sequence. The  
true left end of clone RP5-859D4 is at 119691 in this sequence. The  
true right end of clone RP11-199014 is at 19977 in this sequence.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. RP5-989D7 is from the  
library RPCI-5 constructed by the group of Pieter de Jong. For  
further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2.

## FEATURES

source

Location/Qualifiers  
1..119790

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="20"

/clone="RP5-989D7"

/clone\_lib="RPCI-5"

1..161

/gene="dJ989D7.1"

<1..161

/gene="dJ989D7.1"

/product="dJ989D7.1 (novel transcript)"

/notes="continued from BA199014.1 in Em:AL162504"

match: cDNAs: Em:AY007089"

/evidence=not\_experimental

148..153

/gene="dJ989D7.1"

945..3332

/notes="TIGG81 repeat: matches 1..2418 of consensus"

4130..4386

/note="match: GSS: Em:AQ380734"

4294..4326

/notes="LIMB8 repeat: matches 5834..5865 of consensus"

7201..8108

/notes="LIM4 repeat: matches 4784..5705 of consensus"

complement(8165..8633)

/note="match: GSS: Em:AQ277139"

8629..10044

/notes="LIM4 repeat: matches -260..1191 of consensus"

10045..10190

/notes="73 copies 2 mer tt 57% conserved"

10382..10435

/note="27 copies 2 mer tt 74% conserved"

11843..12150

/notes="AluX repeat: matches 1..308 of consensus"

repeat\_region

12350..12538

/note="MIR repeat: matches 60..249 of consensus"

repeat\_region

12659..13160

/note="MLT1J repeat: matches 13..513 of consensus"

repeat\_region

14687..14768

/note="HAI1 repeat: matches 271..360 of consensus"

repeat\_region

15106..15348

/note="MER2 repeat: matches 1..345 of consensus"

misc\_feature

15368..15768

/note="match: GSS: Em:AQ772240"

misc\_feature

15375..15560

/note="match: GSS: Em:AQ012918"

repeat\_region

15807..16507

/note="LIMB2 repeat: matches 5469..6165 of consensus"

repeat\_region

16561..17220

/note="LIMB2 repeat: matches 4871..5474 of consensus"

repeat\_region

17218..17509

/note="LIMB8 repeat: matches 5848..6168 of consensus"

repeat\_region

17581..17733

/note="MER8 repeat: matches 87..239 of consensus"

repeat\_region

17734..17987

/note="LIMB1 repeat: matches 5782..6043 of consensus"

repeat\_region

18038..18183

/note="LIMB10 repeat: matches 6174..6316 of consensus"

repeat\_region

18188..18525

/note="LIMB7 repeat: matches 2409..2740 of consensus"

repeat\_region

18526..18878

/note="THE1A repeat: matches 1..354 of consensus"

repeat\_region

18879..22631

/note="LIMB7 repeat: matches 2740..6162 of consensus"

repeat\_region

23729..24007

/note="LIMB1 repeat: matches 5598..5892 of consensus"

repeat\_region

24031..24127

/note="LIMB10 repeat: matches 5993..6095 of consensus"

repeat\_region

24128..24439

/note="MER7A repeat: matches 1..345 of consensus"

repeat\_region

24440..24524

/note="LIMB10 repeat: matches 6095..6317 of consensus"

repeat\_region

25218..27841

/note="LIMC1 repeat: matches 3550..6326 of consensus"

misc\_feature

complement(27586..28074)

repeat\_region

28283..28491

/note="match: GSS: Em:AQ605230"

repeat\_region

28592..28729

/note="LTR16A repeat: matches 148..378 of consensus"

repeat\_region

29416..29585

/note="MIR repeat: matches 26..167 of consensus"

repeat\_region

30385..30489

/note="MER5B repeat: matches 1..177 of consensus"

repeat\_region

30759..31042

/note="MIR repeat: matches 40..153 of consensus"

repeat\_region

31306..31349

/note="MER33 repeat: matches 47..323 of consensus"

repeat\_region

31491..31825

/note="MER77 repeat: matches 38..400 of consensus"

misc\_feature

complement(31606..32269)

repeat\_region

32371..32509

/note="match: GSS: Em:AQ266996"

repeat\_region

32510..33153

/note="LTR16C repeat: matches 81..221 of consensus"

repeat\_region

33154..33323

/note="MER51B repeat: matches 1..617 of consensus"

misc\_feature

35252..35712

/note="match: GSS: Em:AQ700733"

repeat\_region

35693..36037

/note="match: STS: Em:L16414"

repeat\_region

35731..35802

/note="MLT1J repeat: matches 124..196 of consensus"

misc\_feature

35934..36427

/note="match: GSS: Em:AQ429711"

repeat\_region

35967..36020







REFERENCE  
AUTHORS

Rattus.  
1 (bases 1 to 343188)  
Muzny, D., Marie, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, P., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwunonu, G., Olarunpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, E., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Streimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

TITLE  
JOURNAL

REFERENCE  
2 (bases 1 to 343188)  
Worley, K.C.

TITLE  
JOURNAL

Submitted (12-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 343188)

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
Rat Genome Sequencing Consortium.

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23816565

## COMMENT

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GSNR  
Center clone name: CH230-271E17  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 265870 bases at least Q40  
Consensus quality: 268598 bases at least Q30  
Consensus quality: 270491 bases at least Q20  
Estimated insert size: 281218; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation table.

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 29799: contig of 29799 bp in length  
29800 29899: gap of unknown length  
29900 139979: contig of 110080 bp in length  
139980 140079: gap of unknown length  
140080 153699: contig of 13620 bp in length  
153700 153799: gap of unknown length  
153800 333380: contig of 179581 bp in length  
333381 333480: gap of unknown length  
333481 334536: contig of 1056 bp in length  
334537 334637: gap of unknown length  
334637 335858: contig of 1222 bp in length  
335859 337051: contig of unknown length  
337052 337151: gap of unknown length  
337152 338449: contig of 1298 bp in length  
338450 338550: gap of unknown length  
338550 339794: contig of 1245 bp in length  
339795 339894: gap of unknown length  
339895 343188: contig of 3294 bp in length.

FEATURES  
source

1. 343188  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-271E17"  
4552. .5701  
/note="wgs\_end\_extension  
clone\_end:T7"  
complement(17244. .17621)  
/note="clone\_boundary  
clone\_end:T7  
site:  
end\_sequence:RXAF333TUB"  
17949. .19643  
/note="wgs\_contig"  
20286. .21465  
/note="wgs\_contig"  
140080. .143453  
/note="wgs\_contig"  
14497. .153699  
/note="wgs\_contig"

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

misc\_feature 153800..179198  
/note="wgs\_contig"  
177457..178312  
/note="clone\_boundary"  
clone\_end:Sp6  
site:  
end\_sequence:BZ218560"  
179249..180714  
/note="wgs\_end\_extension"  
clone\_end:Sp6"

ORIGIN

Query March 9.2%; Score 43; DB 2; Length 343188;  
Best Local Similarity 59.3%; Pred. No. 2.6;  
Matches 73; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2 AATGTTGGCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGAAGCTGCTGC 61  
|||||  
Db 135829 AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 135770  
|||||

QY 62 TGTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGTGTGCGTGCCCAAG 121  
Db 135769 TGTGCTGCTGCTGCTGCTGCTTATCTGTTTTTCCATTGAAGTGTGGAAGGCCAAGAT 135710  
|||||

QY 122 TCT 124  
|||

Db 135709 TCT 135707

RESULT 38

AC096041

LOCUS Rattus norvegicus clone CH230-15A4, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
4 unordered pieces.  
AC096041

AC096041.6 GI:24941294  
HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_ENRICHED.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
1 (bases 1 to 224334)  
Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blau, P., Burrell, K., Blyth, P., Brown, M.,  
Bryant, N., Buha, C., Burch, P., Burrell, K., Ceasar, H., Chen, Y., Chen, Z., Chu, J.,  
Cardenas, J., Carter, K., Cavazos, I., Ceasar, H., Chen, Y., Chen, Z., Chu, J.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabis, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S., L., Hodgson, A., Hogue, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Lit, Y., London, P., Longacre, S., Lopez, J.,  
Lorensuewa, L., Loulseghe, H., Lozano, R., Lu, X., Ma, J.,  
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,



```

/translacion="MSNTVLADSWGTLLSWTVFLLGTSSADSGVWQSPRRHIIKEX
GGRVLTCIPISGNSVWYQTLGKELKFLIQHYEKVERDKGFLPSRFSVQOPDDYH
SEMMSALELEDSAMYFCASLGTGGRHPVLWARSAPRV"
join(401..479,609..616)
/gene="Tcrb"
617..>955
/gene="Tcrb"
/product="T cell receptor beta-chain"
join(401..479,609..616)
/gene="Tcrb"
<401..479
/gene="Tcrb"
/number=1
/number=1
480..608
/gene="Tcrb"
/number=1
609..957
/gene="Tcrb"
/number=2
/number=2
901..902
/organism="Mus musculus"
911..912
/organism="Mus musculus"
/organism="Mus musculus"
102 bp upstream of XbaI site.

Query Match          9.1%; Score 42.6; DB 10; Length 982;
Best Local Similarity 78.5%; Pred. No. 2.7;
Matches 51; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 17 ACCTCCCCCAAGCCCTTACCTATGCTGTGCTAAGCTGCTGCTGCTGCTGCTG 76
Db 502 ACTGCACCTAAGTCCCTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 561

QY 77 TTAAA 81
Db 562 ATAGA 566

RESULT 41
DROZSTA/c
LOCUS
DEFINITION
Drosophila virilis zeste gene, complete cds.
ACCESSION
M76700
VERSION
M76700.1 GI:158825
KEYWORDS
DNA-binding; chromosomal localization; gene regulation;
transvection; zeste protein.
SOURCE
Drosophila virilis
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
Chen,J.D., Chan,C.S. and Pirrotta,V.
Conserved DNA binding and self-association domains of the
Drosophila zeste protein
Mol. Cell. Biol. 12 (2), 598-608 (1992)
J2123185
MEDLINE
PUBMED
1732733
COMMENT
Original source text: Drosophila virilis DNA.
FEATURES
Location/Qualifiers
1..2708
/organism="Drosophila virilis"
/mol_type="genomic DNA"
/db_xref="taxon:7244"
/map="chromosome X"
/germline
join(<594..1015,1096..1658,1747..2618)
/gene="zeste"
join(<594..1015,1096..1658,1747..2618)
/gene="zeste"
/codon_start=1
/product="zeste"

sig_peptide
mat_peptide
gene
exon
intron
misc_recomb
misc_recomb
ORIGIN
102 bp upstream of XbaI site.

Query Match          9.1%; Score 42.6; DB 3; Length 2708;
Best Local Similarity 56.9%; Pred. No. 2.8;
Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 6 TTGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTGCTGCTGCTGCTGCTGCT 65
Db 1271 TTGATACCTCGGCTCCACCTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1212

QY 66 GCTGCTGCTGCTTAAAGGCTCATGCTTGAAGTGGGACTGCTGCTGCTGCTGCTGCT 125
Db 1211 GTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1152

QY 126 TTCTGCCACTGACGCC 142
Db 1151 GGCCTGCGCTGCCACCC 1135

RESULT 42
AL603889/c
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-407M20 on chromosome 11,
complete sequence.
ACCESSION
AL603889
VERSION
AL603889.5 GI:20068524
KEYWORDS
HTG.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1
AUTHORS
Van Hellmond,Z.
TITLE
Direct Submission
JOURNAL
Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19031503.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all

```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, ENBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-407M20 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

```

FEATURES             source
    Location/Qualifiers
        1..145052
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /chromosome="11"
            /clone="RP23-407M20"
            /clone_lib="RPCI-23"

ORIGIN
Query Match          9.1%; Score 42.6; DB 10; Length 145052;
Best Local Similarity 78.5%; Pred No. 3.3;
Matches 51; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 13 ATCCACCTCCCGCCAGCCCTTACCTGATGCTGCTGCTGCTGCTGCTGCTGCTG 72
   || || || || || || || || || || || || || || || || || || || ||
Db 129089 ATACACCAACCCCTCTGCGCTTAAATATGCTGCTGCTGCTGCTGCTGCTG 129030

Qy 73 CTGCT 77
   || || || || || || || || || || || || || || || || || || || ||
Db 129029 CTGCT 129025

RESULT 43
AC128156 LOCUS
DEFINITION Rattus norvegicus clone CH230-287H17, WORKING DRAFT SEQUENCE.
ACCESSION AC128156
VERSION AC128156.3 GI:25007826
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 199221)
Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blythe,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos J., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flegg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.W., Gabist,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebreygeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

```

```

Kovis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Loulsegad,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montenayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelamh,O., Okwuonu,G., Olarunpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu.L.L.,
Puzos,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rockey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snajds,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Soralle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 199221)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 199221)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23908235.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZOO
Center clone name: CH230-287H17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 186979 bases at least Q40
Consensus quality: 189196 bases at least Q30
Consensus quality: 190602 bases at least Q20
Estimated insert size: 190707; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently

```

\* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 199221: contig of 199221 bp in length.

## FEATURES

source  
 1. 199221  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-28/H17"  
 1. 1558  
 /note="wgs end extension  
 clone\_end:T7"  
 4765..6815  
 /note="wgs end extension  
 clone\_end:T7"  
 7600..8500  
 /note="clone boundary  
 clone\_end:T7"  
 site:  
 end\_sequence:BZ134792"  
 192560..193460  
 /note="clone boundary  
 clone\_end:Sp6  
 site:  
 end\_sequence:BZ134793"  
 195186..196606  
 /note="wgs end extension  
 clone\_end:Sp6"  
 197746..199221  
 /note="wgs end extension  
 clone\_end:Sp6"

## ORIGIN

Query Match 9.1%; Score 42.6; DB 2; Length 199221;  
 Best Local Similarity 58.1%; Pred. No. 3.3;  
 Matches 75; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
 QY 4 TGTGTGCTGCTACCTCCCAAGCCCTTACCTATGCTGCTGCTAAGCTGCTGCTG 63  
 Db 16884 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 16943  
 QY 64 CTGCTGCTGCTCTTAAAGCTCATGCTTGGAGTGGGACTGCTGCTGCCAGAAAGTC 123  
 Db 16944 CTGCTGCTGCTCTTCTCTAGCTAGTGGTCAGCAGAGCTTTCTTGACTCTATACAAA 17003  
 QY 124 TCTTCTGCC 132  
 Db 17004 GCTGCTCC 17012

RESULT 44  
 MMAE000663  
 LOCUS MMAE000663 250611 bp DNA linear ROD 04-OCT-2002  
 DEFINITION Mus musculus TCR beta locus from bases 1 to 250611 (section 1 of 3)  
 of the complete sequence.  
 ACCESSION AE000663 AE000522  
 VERSION AE000663.1 GI:2358069  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 250611)  
 AUTHORS Chen, F., Rowen, L., Hood, L. and Rothenberg, E.V.  
 TITLE Differential transcriptional regulation of individual TCR V beta  
 segments before gene rearrangement  
 JOURNAL J. Immunol. 166 (3), 1771-1780 (2001)

MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

21103195  
 11160223  
 2 (bases 1 to 250611)  
 Rowen, L., Koop, B.F., Boyesen, C., Wang, K., Ahearn, M.E., Qin, S.,  
 Lee, I., Seto, J., Acharya, C., Ankner, M., Alving, W., Chen, L.,  
 Paepker, B., Baskin, D., Jerome, N., Swartzell, S., Gilbert, T.,  
 Faust, J., Loretz, C., Bumgarner, R. and Hood, L.  
 Sequence of the mouse T cell receptor locus  
 Unpublished

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

3 (bases 1 to 250611)  
 Rowen, L., Smit, A.F.A. and Hood, L.  
 Comparison of the human and mouse T cell receptor beta and  
 tryptsinogen loci  
 Unpublished

JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

4 (bases 1 to 250611)  
 Rowen, L.  
 Direct Submission  
 Submitted (20-JUL-1997) Department of Molecular Biotechnology, Box  
 357730 University of Washington, Seattle, Washington 98195, USA  
 Sequencing methodology: high redundancy shotgun. Based on  
 overlapping cosmids, the precision of this sequence is

## COMMENT

99.93. Libraries were derived from two strains but there was no  
 discernable sequence variation between them. This sequence  
 contains numerous internal duplications, especially in the  
 tryptsinogen genes. Interspersed Repeats were identified with  
 RepeatMasker (available from  
 http://ftp.genome.washington.edu/RX/RepeatMasker.html) Several new  
 repeats were identified by Arian Smit, and these have not yet been  
 fully classified. Simple sequence repeats were identified with  
 sputnik (available from  
 http://serac.mbt.washington.edu/~chrisa/software/sputnik.html)

The authors thank Christopher Nelson for his gift of cosmids from  
 Dennis Loh's library.

## FEATURES

Location/Qualifiers  
 1. 250611  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="6"  
 1. 36170  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /clone="cosmid D23"  
 /clone\_lib="Dennis Loh"  
 34707..76259  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="129"  
 /db\_xref="taxon:10090"  
 /clone="cosmid M4B2"  
 /clone\_lib="subcloned from BAC 57B22, Genome Systems  
 library"  
 63060..97731  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /clone="cosmid 0742C"  
 /clone\_lib="Kai Wang"  
 96883..131111  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /clone="cosmid 79A"  
 110337..144911  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="cosmid 74A"

## source

## source

## source

## source

## source





```

REFERENCE
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (10-JAN-2003) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA
              6 (bases 1 to 139786)
              Wilson,R.
REFERENCE
AUTHORS      Direct Submission
TITLE        Submitted (13-NOV-2003) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On Jan 10, 2003 this sequence version replaced gi:24211432.
              ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu
              Contact: submission@watson.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: M_BB0445C12
              -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1..139786
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="3"
/map="3"
/clone="RP24-445C12"
/clone_lib="RPCI-24"
1..91
/rpt_family="L1"
repeat_region 845..887
/rpt_family="L1"
repeat_region 995..1167
/rpt_family="L1"
repeat_region 1231..1284
/rpt_family="ERV1"
repeat_region 1454..1600
/rpt_family="L1"
repeat_region 2273..2774
/rpt_family="RMR15"
repeat_region 4149..4592
/rpt_family="L1"
repeat_region 7500..7532
/rpt_family="L1"
repeat_region 7533..7724

```

[illegible]







## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23322236.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWTC  
Center clone name: CH230-436I19  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 206076 bases at least Q40  
Consensus quality: 209800 bases at least Q30  
Consensus quality: 212212 bases at least Q20  
Estimated insert size: 201535; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 4140: contig of 4140 bp in length  
\* 4141 4240: gap of unknown length  
\* 4241 14717: contig of 10477 bp in length  
\* 14718 14817: gap of unknown length  
\* 14818 202669: contig of 187852 bp in length  
\* 202670 202769: gap of unknown length  
\* 202770 218085: contig of 15316 bp in length  
\* 218086 218185: gap of unknown length  
\* 218186 223017: contig of 4832 bp in length  
\* 223018 223117: gap of unknown length  
\* 223118 224236: contig of 1119 bp in length  
\* 224237 224336: gap of unknown length  
\* 224337 225470: contig of 1134 bp in length  
\* 225471 225570: gap of unknown length  
\* 225571 227570: contig of 2000 bp in length  
\* 227571 227670: gap of unknown length  
\* 227671 228961: contig of 1291 bp in length  
\* 228962 229061: gap of unknown length  
\* 229062 230156: contig of 1095 bp in length  
\* 230157 230256: gap of unknown length  
\* 230257 231359: contig of 1103 bp in length  
\* 231360 231459: gap of unknown length  
\* 231460 232804: contig of 1345 bp in length  
\* 232805 232904: gap of unknown length  
\* 232905 234228: contig of 1324 bp in length.

FEATURES  
Source

1. 234228  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-436I19"  
misc\_feature 1..1386  
/note="wgs\_contig"

misc\_feature 14818..16195  
/note="wgs\_contig"  
misc\_feature 200359..202669  
/note="wgs\_contig"  
misc\_feature 205939..207188  
/note="wgs\_contig"  
misc\_feature 214714..215721  
/note="wgs\_contig"  
misc\_feature 218186..220168  
/note="wgs\_contig"  
ORIGIN  
Query Match 9.1%; Score 42.4; DB 2; Length 234228;  
Best Local Similarity 76.5%; Pred.No.3.7;  
Matches 52; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 10 CCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTG 69  
Db 86795 CTTCTTCCCTCCCTCCCTCCCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTG 86736  
QY 70 CTGCTGCT 77  
Db 86735 CTGCTGCT 86728

Search completed: May 25, 2004, 16:30:27  
Job time : 3370 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 15:09:31 ; Search time 2857 Seconds  
(without alignments)  
4881.217 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED  
Perfect score: 467  
Sequence: 1 caatgttgcctaccct.....taataaacagttaaaagctg 467

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_estnum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estic:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	466	99.8	728	12	BM975075 UI-CP-EC1
C 2	462.2	99.0	555	10	BE855577
C 3	456	97.6	482	9	AI871469
C 4	454	97.2	517	14	CA444588



```

78 47.2 10.1 1203 29 CNS015WU AL106008 Drosophil
79 47.8 10.1 1299 10 AW485179
80 46.8 10.0 1043 29 CNS0145P AL103735 Drosophil
81 46.6 10.0 1201 13 BX374796 BX374796
82 46.2 9.9 394 13 BY086456 BY086456
83 46.2 9.9 693 13 BX381320 BX381320
84 46.2 9.9 997 29 CNS003AV AL064073 Drosophil
85 46 9.9 843 29 CNS000CS1 AL059666 Drosophil
86 45.8 9.8 462 9 AA392984 LD12163.5
87 45 9.6 896 29 CNS000P8 AL057462 Drosophil
88 45 9.6 949 29 CNS022AU AL220719 Tetraodon
89 45 9.6 1201 9 AL513975 AL513975
90 44.6 9.6 672 28 BH396372 BH396372 AG-ND-159
91 44.4 9.5 844 29 CNS0052P AL056652 Drosophil
92 44.2 9.5 660 28 BZ447727 BZ447727 BONOT32TR
93 43.8 9.4 509 14 CD342739 CD342739 EtESTes56
94 43.8 9.4 578 14 CD659850 CD659850 EtESTef37
95 43.8 9.4 778 29 CNS010UP AL099451 Drosophil
96 43.8 9.4 1014 29 CNS0071P AL066670 Drosophil
97 43.8 9.4 1103 13 BX334917 BX334917
98 43.6 9.3 586 14 CD660017 CD660017 EtESTef38
99 43.2 9.3 381 13 BX360796 BX360796
100 43.2 9.3 728 29 CE142464 CE142464 tigr-gss-

```

## ALIGNMENTS

```

RESULT 1
LOCUS BM975075/c
DEFINITION UI-CF-EC1-acf-1-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
ACCESSION BM975075
VERSION BM975075.1 GI:19592666
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 728)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cDNA
sequence: 408-444. >(CAC)n$Simple_repeat 685-728,
>(GGCA)n$Simple_repeat (matched complement)
Seq primer: M13_FORWARD
POLYA=Yes.
Location/Qualifiers
1..728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-acf-1-23-0-UI"
/tissue_type="Lung"

```

## FEATURES

source

```

/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EC1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGTCTTAC.
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_LIB=UI-CF-EC1
TAG_SEQ=AAGTGTCTAC"

```

## ORIGIN

```

Query Match 99.8%; Score 466; DB 12; Length 728;
Best Local Similarity 100.0%; Pred. No. 3.6e-111;
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTTGCCTATCCACTCCCTCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 60
Db |||||
484 CAATGTTTGCCTATCCACTCCCTCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTG 425
QY 61 CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGAGTGGGACTGCTGGTCCCGAGAAA 120
Db |||||
424 CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGAGTGGGACTGCTGGTCCCGAGAAA 365
QY 121 GTCCTTTTGGCACTGACGCGCCCATCAGGATGGGCTTCTTTCCCTTCCTTCCTG 180
Db |||||
364 GTCCTTTTGGCACTGACGCGCCCATCAGGATGGGCTTCTTTCCCTTCCTTCCTG 305
QY 181 TGCTCCCTGCTCATCGGCTGCCATGACCTGCAGCAGCCAGCCCGCGTGGGAGGG 240
Db |||||
304 TGCTCCCTGCTCATCGGCTGCCATGACCTGCAGCAGCCAGCCCGCGTGGGAGGG 245
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 300
Db |||||
244 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 185
QY 301 GGCTAGGGGGCTGGCTTTATTAAGTGGTGTGTTATGATCTTATATACTAATTTATACAA 360
Db |||||
184 GGCTAGGGGGCTGGCTTTATTAAGTGGTGTGTTATGATCTTATATACTAATTTATACAA 125
QY 361 AGATATTAAAGCCCTGTTTCATTAAAGAAATCTCCCTTCCCTGCTGTTCAATGTTGTAA 420
Db |||||
124 AGATATTAAAGCCCTGTTTCATTAAAGAAATCTCCCTTCCCTGCTGTTCAATGTTGTAA 65
QY 421 AGATTTGTTCTGTGTAATAATGCTTTTATAATAAACAAGTTAAAGCT 466
Db |||||
64 AGATTTGTTCTGTGTAATAATGCTTTTATAATAAACAAGTTAAAGCT 19

```

## RESULT 2

```

BE855577/c
LOCUS BE855577
DEFINITION 7g10e01.x1 NCI CGAP Brn23 Homo sapiens cDNA clone IMAGE:3306072 3,
similar to contains element OFR repetitive element ;, mRNA
sequence.
ACCESSION BE855577
VERSION BE855577.1 GI:10367753
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 555)  
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1998)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-x@mail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL, send email to:  
infoimage.lnl.gov  
Seq primer: -40UP from Gibco  
High quality sequence stop: 472.  
Location/Qualifiers  
source  
1..555  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3306072"  
/tissue\_type="glioblastoma (pooled)"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Brn23"  
/note="Organ: brain; Vector: pTZ19+ (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TGTTCAACTTCGAAGTGGCGGCCATACCTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pTZ19+ vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Fatima Bonaldo."

ORIGIN

Query Match	99.0%	Score	462.2;	DB 10;	Length	555;	
Best Local Similarity	99.4%;	Pred.	No. 3.2e-110;	Indels	0;	Gaps	0;
Matches	464;	Conservative	0;	Mismatches	3;		

```

QY      1 CAATGTTTGCTATCCACCTCCCCCAGCCCTTTTACCTATGCTGTCTAAGCGTGCCTG 60
Db      468 CAATGTTTGCTATCCACCTCCCCCAGCCCTTTTACCTATGCTGTCTAAGCGTGCCTG 409

QY      61 CTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGACATGTCGGTGCCTCCAGAAA 120
Db      408 CTGCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGACATGTCGGTGCCTCCAGAAA 349

QY      121 GTCTCTTCTGCCACTGACGCCGCCCATCAGGATTTGGCCTTCTTCCCCTTCTCTTCTG 180
Db      348 GTCTCTTCTGCCACTGACGCCGCCCATCAGGATTTGGCCTTCTTCCCCTTCTCTTCTG 289

QY      181 TGCTCTCTGCTCATTCGGCTGCTCATGACCTGTCAGCCAGCCAGCCCGCTGGGGAAGGG 240
Db      288 TGCTCTCTGCTCATTCGGCTGCTCATGACCTGTCAGCCAGCCAGCCCGCTGGGGAAGGG 229

QY      241 GAGAAAGTGGGGATGGCTAAGAAGCTGGGAGATAGGGAACAAGAAGGCTAGTGGGTG 300
Db      228 GAGAAAGTGGGGATGGCTAAGAAGCTGGGAGATAGGGAACAAGAAGGCTAGTGGGTG 169

QY      301 GGCTAGGGGGCTGCCCTATTATAAGTGGTGTATTATGATCTTATCTAATTTATCAA 360
Db      168 GGCTAGGGGGCTGCCCTATTATAAGTGGTGTATTATGATCTTATCTAATTTATCAA 109

QY      361 AGATATTAAAGCCCTGTTTCATTAAAGATTTGTTCCCTTCCCTGCTGTTCAATGTTGTAA 420
Db      108 AGATATTAAAGCCCTGTTTCATTAAATTAATTTGTTCCCTTCCCTGCTGTTCAATGTTGTAA 49

QY      421 AGATTGTTCTGTGTAAATATGCTCTTTATAATAAACACAGTTAAAGAGCTG 467

```

Db 48 AGATTGTTCTGCTAAATATGCTTTTATAATAACAGTTAAAGAGCTG 2

RESULT 3  
AI871469/c

LOCUS  
DEFINITION  
w167409.x1 NCI\_CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2429969 3,  
similar to contains element TAR1 repetitive element i, mRNA  
sequence.

ACCESSION  
AI871469

VERSION  
AI871469.1

KEYWORDS  
GI:5545518

SOURCE  
EST.

ORGANISM  
Homo sapiens (human)

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS  
1 (bases 1 to 482)

TITLE  
NCI/MINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTCAP), Tumor Gene Index

JOURNAL  
Unpublished (1998)

COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-femail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert Length: 1409 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 450.  
Location/Qualifiers  
1. .482  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2429969"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Brn25"  
/notes="Organ: Brain; Vector: pVT73D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo (dT) primer [5',  
TGTTACCAATCGAGTGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTTTTTT  
T 3'], double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pVT3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 97.6%; Score 456; DB 9; Length 482;  
Best Local Similarity 99.8%; Pred.No. 1.3e-108;  
Matches 467; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CAATGTTTCCTATCACTCCCCCAAGCCCTTTACCTGCTGCTGCTAAAGCTGCG 60  
|||||

Db 468 CAAATGTTGCCTATCACTCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTG 409  
|||||

QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGCTGCTCGGTGCCAGAAA 120  
|||||

Db 408 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGAGCTGCTCGGTGCCAGAAA 349  
|||||

QY 121 GTCTCTTTGCCACTGACGCCGCCCAATCAGGAGTTGGGCTTTCTTCCCGCTTCCTTCTG 180  
|||||

Db 348 GTCTCTTCTGCCACTGACGCCGCCCAATCAGGAGTTGGGCTTTCTTCCCGCTTCCTTCTG 289  
|||||

QY 181 TGCTCTCTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGCTGGGGAAGGG 240  
 Db 288 TGCTCTCTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGCTGGGGAAGGG 229  
 QY 241 GAG-AAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGT 299  
 Db 228 GAGAAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGT 169  
 QY 300 GGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTTATACA 359  
 Db 168 GGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTTATACA 109  
 QY 360 AAGATATTAAGGCGCTGTTCAATTAAGAAATGTTCCCTCCCTGGTGTCAATGTTTGTGA 419  
 Db 108 AAGATATTAAGGCGCTGTTCAATTAAGAAATGTTCCCTCCCTGGTGTCAATGTTTGTGA 49  
 QY 420 AAGATTTCTGTGTAATATGCTTTATTAATAACAGTTAAAGCTG 467  
 Db 48 AAGATTTCTGTGTAATATGCTTTATTAATAACAGTTAAAGCTG 1

RESULT 4  
 CA444588/c  
 LOCUS CA444588  
 DEFINITION UI-H-DTI-awl-m-06-0-UI.s1 NCI CGAP\_DTI Homo sapiens CDNA clone  
 ACCESSION CA444588  
 VERSION CA444588.1 GI:24809008  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 517)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 JOURNAL Tumor Gene Index  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Dr. Jose Mercuende  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing: Dr. M. Bento Soares, University of Iowa  
 Cloning: Dr. M. Bento Soares, University of Iowa  
 Cloning Distribution: Cloning information can be obtained  
 from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
 The following repetitive elements were found in this CDNA  
 sequence: 409-448, >(CAG)n#Simple\_repeat  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

Location/Qualifiers  
 1..517  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-DTI-awl-m-06-0-UI"  
 /issue\_type="Metastatic Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI-CGAP\_DTI"  
 /notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 NCI CGAP\_DTI is a normalized CDNA library containing the  
 following tissue(s): Metastatic Chondrosarcoma in Lung.  
 The library was constructed according to Ronaldo, Lennon  
 and Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an EcoR I adaptor, digested with Not I, and cloned  
 directionally into pT73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dT)18 tail. The sequence tag for this

library is AACTGTTCCG.  
 TAG\_TISSUE=lung metastatic chondrosarcoma  
 TAG\_LIB=UI-H-DTI  
 TAG\_SEQ=AACTGTTCCG"

## ORIGIN

Query Match 97.2%; Score 454; DB 14; Length 517;  
 Best Local Similarity 99.4%; Pred. No. 4.5e-108;  
 Matches 467; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
 QY 1 CAATGTTTCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAAAC---GCTG 57  
 Db 488 CAATGTTTCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTGCTGCTGCTG 429  
 QY 58 CTGCTCTCTGCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGGTCGGTGCCAG 117  
 Db 428 CTGCTCTCTGCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACTGGTCGGTGCCAG 369  
 QY 118 AAGTCTCTTCTGCCACTGACGCCCCCATCAGGATTGGGCTTCTTTCCTCCCTTCTTT 177  
 Db 368 AAGTCTCTTCTGCCACTGACGCCCCCATCAGGATTGGGCTTCTTTCCTCCCTTCTTT 309  
 QY 178 CTGCTCTCTCTCTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237  
 Db 308 CTGCTCTCTCTCTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249  
 QY 238 GGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGG 297  
 Db 248 GGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGG 189  
 QY 298 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTTATA 357  
 Db 188 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTTCTTACTAATTTATA 129  
 QY 358 CAAAGATATTAAGGCGCTGCTTATTAAGAAATGTTTCCCTCCCTGCTGCTCAATGTTTG 417  
 Db 128 CAAAGATATTAAGGCGCTGCTTATTAAGAAATGTTTCCCTCCCTGCTGCTCAATGTTTG 69  
 QY 418 TAAAGATTTCTGTGTAATATGCTTTTATAAACAAGTTAAAGCTG 467  
 Db 68 TAAAGATTTCTGTGTAATATGCTTTTATAAACAAGTTAAAGCTG 19

## RESULT 5

BQ006616/c  
 LOCUS BQ006616  
 DEFINITION UI-H-Ell-aza-n-02-0-UI.s1 NCI CGAP\_Ell Homo sapiens CDNA clone  
 IMAGE:5846209 3', mRNA sequence.  
 ACCESSION BQ006616  
 VERSION BQ006616.1 GI:19731516  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 668)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 JOURNAL Tumor Gene Index  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Dr. Jose Mercuende  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing: Dr. M. Bento Soares, University of Iowa  
 Cloning Distribution: Cloning information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 The following repetitive elements were found in this CDNA  
 sequence: 409-448, >(CAG)n#Simple\_repeat  
 Seq primer: M13 FORWARD  
 POLYA=Yes.  
 Location/Qualifiers



```

Db      90  TTAAGAAATGTTCCCTCCCTGCTGCTCAATGTTTGTAAAGATTGCTCTGTGTAATAT 31
QY      441  GTCTTTATATAAACAGTTAAAGCTG 467
Db      30  GTCTTTATATAAACAGTTAAAGCTG 4

RESULT 7
CA438256/c
LOCUS      CA438256      687 bp      mRNA      linear      EST 08-NOV-2002
DEFINITION UI-H-DT1-aww-g-05-0-UI.s1 NCI_CGAP_DT1 Homo sapiens cDNA clone
            UI-H-DT1-aww-g-05-0-UI 3', mRNA sequence.
ACCESSION  CA438256
VERSION     1
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 687)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TUMOR GENE INDEX
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            The following repetitive elements were found in this cDNA
            sequence: 409-448. >(CAG)n#simple_repeat
            Seq primer: M13 FORWARD
            POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..687
     mol_type="mRNA" /organism="Homo sapiens"
     db_xref="taxon:9606" /mol_type="mRNA"
     clone="UI-H-DT1-aww-g-05-0-UI" /db_xref="taxon:9606"
     tissue_type="Metastatic Chondrosarcoma" /clone="UI-H-DT1-aww-g-05-0-UI"
     dev_stage="Adult" /tissue_type="Metastatic Chondrosarcoma"
     lab_host="DH10B (Life Technologies)" /dev_stage="Adult"
     clone_lib="NCI CGAP DT1" /lab_host="DH10B (Life Technologies)"
     note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
     modified polylinker; Site_1: EcoR I; Site_2: Not I;
     NCI_CGAP_DT1 is a normalized cDNA library containing the
     following tissue(s): Metastatic Chondrosarcoma in Lung.
     The library was constructed according to Bonaldo, Lennon
     and Soares, Genome Research, 6:791-806, 1996. First strand
     cDNA synthesis was primed with an oligo-dT primer
     containing a Not I site. Double stranded cDNA was ligated
     to an EcoR I adaptor, digested with Not I, and cloned
     directionally into pT73-Pac vector. The oligonucleotide
     used to prime the synthesis of first-strand cDNA contains
     a library tag sequence that is located between the Not I
     site and the (dT)18 tail. The sequence tag for this
     library is AACTGTTCCG.
     TAG_TISSUE=lung metastatic chondrosarcoma
     TAG_LIB=UI-H-DT1
     TAG_SEQ=AACTGTTCCG"

ORIGIN
Query Match      94.3%; Score 440.4; DB 14; Length 687;
Best Local Similarity 98.9%; Pred. No. 1.8e-104;
Matches 465; Conservative 0; Mismatches 1; Indels 4; Gaps 2;

QY      1  CAATGTTTCCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAAC---GCTG 57
Db      487  CAATGTTTCCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 429

```

```

QY      58  CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGCCAG 117
Db      428  CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGCCAG 369
QY      118  AAAGTCTCTTCTGCTGCTCAACGCGCCCATCAGGGATTGGGCTTCTTCCCTCTCTTT 177
Db      368  AAAGTCTCTTCTGCTGCTCAACGCGCCCATCAGGGATTGGGCTTCTTCCCTCTCTTT 309
QY      178  CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237
Db      308  CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
QY      238  GGGGAGAAAGTGGGGGATGCTTAAAGAAAGCTGGAGATAGGAAACAGAGAGAGGTAGTGG 297
Db      248  GGGGAGAAAGTGGGGGATGCTTAAAGAAAGCTGGAGATAGGAAACAGAGAGAGGTAGTGG 189
QY      298  GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTCTTATCTAATTTATA 357
Db      188  GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATTCTTATCTAATTTATA 129
QY      358  CAAAGATATTAAAGCCCTGCTTCAATTAAGAAATGTTCCCTGCTGCTGCTGCTGCTG 417
Db      128  CAAAGATATTAAAGCCCTGCTTCAATTAAGAAATGTTCCCTGCTGCTGCTGCTGCTG 69
QY      418  TAAAGATTGTTCTGTTAAATATGCTTTTATATAACAGATTAAAGCTG 467
Db      68  TAAAGATTGTTCTGTTAAATATGCTTTTATATAACAGATTAAAGCTG 19

RESULT 8
BX434223/c
LOCUS      BX434223      889 bp      mRNA      linear      EST 15-MAY-2003
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION  BX434223
VERSION     BX434223.1
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 889)
AUTHORS    Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 10757.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAK025CH09NM1&cluster=10757.r. Contact :
            Feng liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAK025CH09NM1.
            Location/Qualifiers
            1..889
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DE009YJ24"
            /tissue_type="PLACENTA"
            /clone_lib="Homo sapiens PLACENTA"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

FEATURES             source

```



COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewatson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert length: 528 Std Error: 0.00 Seq primer: -28M13 rev2 from Amersham.			
FEATURES	Location/Qualifiers 1. .431			
source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:3804224" /db_xref="taxon:9606" /clone="IMAGE:489637" /sex="female" /dev_stage="adult" /lab_host="DH10B" /clone_lib="Scares_pregnant_uterus_NbHPU" /note="Organ: uterus; Vector: pYT3-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AAGTGAAGAATTCGGCGCCGCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pYT3 vector. Library went through one round of normalization. Library constructed by M. Fatima Honaldo."			
ORIGIN	Query Match 90.0%; Score 420.2; DB 9; Length 431; Best Local Similarity 99.3%; Pred. No. 3.1e-99; Matches 422; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	7	TTCCCTATCCACCTCCCCAAGCCCTTTACCTATGCTGCTGTAAACGCTGCTGCTGCTG 66		
Db	6	TTTGCTATCACCTCCCCAAGCCCTTTACCTATGCTGCTGTAAACGCTGCTGCTGCTG 65		
QY	67	CTGCTGCTGCTTAAAGCTCATGCTCGAGTGGGACTGGTCCGTGCCAGAACTCTCT 126		
Db	66	CTGCTGCTGCTTAAAGCTCATGCTCGAGTGGGACTGGTCCGTGCCAGAACTCTCT 125		
QY	127	TCGTGCCACTGACGCGCCCATCAGGGAATGGGACCTCTTTCCCTCCCTTCTGTGCTC 186		
Db	126	TCGTGCCACTGACGCGCCCATCAGGGAATGGGACCTCTTTCCCTCCCTTCTGTGCTC 185		
QY	187	CTGCTCATCGGCTGTCATGACCTCGAGCCAGCCAGCCCGTGGGAGGGGAGAA 246		
Db	186	CTGCTCATCGGCTGTCATGACCTCGAGCCAGCCAGCCCGTGGGAGGGGAGAA 245		
QY	247	GTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTAG 306		
Db	246	GTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTAG 305		
QY	307	GGGGGCTGCCCTATTAAAGTGGTGTATTATGATCTTATCTAATTTATACAAGATAT 366		
Db	306	GGGGGCTGCCCTATTAAAGTGGTGTATTATGATCTTATCTAATTTATACAAGATAT 365		
QY	367	TAAAGGCTGTTCATTAAAGAAATGTTCCCTTCCCTCGTGTTCATGTTCTAAAGATTG 426		
Db	366	TAAAGGCTGTTCATTAAAGAAATGTTCCCTTCCCTCGTGTTCATGTTCTAAAGATTG 425		
QY	427	TTCTG 431		
Db	426	TTCTG 430		
RESULT 11				
AA101878/c				
LOCUS	AA101878	458 bp	mRNA	linear
DEFINITION	zk85f07.s1 Scares pregnant uterus NbHPU Homo sapiens cDNA clone			
EST 11-MAY-1997				

	IMAGE:489637 3', mRNA sequence.
ACCESSION	AA101878
VERSION	AA101878.1 GI:1645281
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 458) Hillier,L., Lemon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubouque,T., Favello,A., Gish,W., Hawkins,J., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE	Generation and analysis of 280,000 human expressed sequence tags
JOURNAL	Genome Res. 6 (9), 807-828 (1996)
MEDLINE	97044478
PUBMED	8889549
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: estewartson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. Insert Length: 528 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham. Location/Qualifiers 1..458 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="GDB:3804224" /db_xref="taxon:9606" /clone="IMAGE:489637" /sex="female" /dev_stage="adult" /lab_host="DH10B" /clone_lib="Soares pregnant uterus NbHpV" /note="Organ: uterus; Vector: pVT73-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dm) primer [5' AATCGAAGAATTCCGGCGGCCCTTTTTCCTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pVT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."
FEATURES	
source	
ORIGIN	
Query Match	89.8%; Score 419.2; DB 9; Length 458;
Best Local Similarity	97.3%; Pred.No.5,7e-99;
Matches	435; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY	22 CCCCAAGCCCTTTACTATGCTGTCTGAACGCTGCTGTGTGTGTGTGTGTGTCTTTAAA 81
Db	458 CCCCAAGCCCTTTACTATGCTGTCTTNAACTGCTGTGTGTGTGTGTGTGTGTATAA 399
QY	82 GGCT-CATGCTTGAGTGGGAGCTGTCGGTGCCTGCCAGAAAAAGTCTCTTTCGCATGACGC 140
Db	398 GGCTCCCATGCTTGAGTGGGAGCTGTTGCTGCTGCCAGAAAAAGTCTCTTCTGCACTGACGC 339
QY	141 CCCCATCAGGGATGGGCGCTTTTCCCCTTCTGTTCTGTTCTGCTGCTCATCGCC 200
Db	338 CCCCATCAGGGATGGGCGCTTTTCCCCCTTCTTTCTGTGCTCTNCTNCTCATPGGCC 279
QY	201 TGCATGACCTGCAGCCCAAGCCAGCCCGCTGGGGAAGGGAGAGAAAGTGGGGATGGCTA 260
Db	278 TGCCATGACCTGCAGCCAGCCCGAGCCCTGGGAAGGGGAGAAAGTGGGGATGGCTA 219
QY	261 AGAAGCTGGGAGATAGGGAACAAGAGAGGGTAGTGGGTGGGCTAGGGGGGTGCCCTTAT 320







```

Db      20 TAAAGATTGADGTGGTAA 2

RESULT 15
LOCUS   BE463932/c
DEFINITION hxb3b11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3194397 3',
mrna sequence.
ACCESSION BE463932
VERSION   BE463932
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 406)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES             source
     1..406
     /organism="Homo sapiens"
     /mol_type="mrna"
     /db_xref="taxon:9606"
     /clone="IMAGE:3194397"
     /lab_host="DH10B"
     /clone_lib="NCI_CGAP Kid11"
     /note="Organ: kidney; Vector: pTV73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
ORIGIN
Query Match      85.1%; Score 397.2; DB 10; Length 406;
Best Local Similarity 99.0%; Pred. No. 3.2e-93;
Matches 399; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      65 TCGTCTGCTGCTTAAAGCTCATGCTTGGAGTGGGACGTGGTGGTGGCCAGAAAGTCT 124
      406 TGCTGCTGNTGCTTAAAGGATCATGCTTGGAGTGGGACGTGGTGGTGGCCAGAAAGTCT 347

Qy      125 CTTCTGCCACTGACGCCCTCATAGGATGGGCTCTTTCCCTTCCTTTCTGTC 184
      346 TTCTGTCACATGACGCCCTCATAGGATGGGCTCTTTCCCTTCCTTTGTC 287

Qy      185 TCCTGCTCATGCGCCTGCCATGACCTGCACGCCAGCCCGCGTGGGAGGGGAGA 244
      286 TCCTGCTCATGCGCCTGCCATGACCTGCACGCCAGCCCGCGTGGGAGGGGAGA 227

Qy      245 AAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAAAGGATAGTGGGTGGGCT 304
      226 AAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAAAGGATAGTGGGTGGGCT 167

Qy      305 AGGGGGCTGCTATTATTAAGTGGTGTATGATTTCTATCTATTATCTATCAAGAT 364

```

```

Db      166 AGGGGGCTGCCTTATTTAAAGTGGTGTATGATTCTTATCTAATTATACAAAGAT 107
      365 ATTAAGGCCCTGTTTCATTAGAAATTTGTCCTTCCCTGCTTCAATGTTGTAAAGAT 424
      106 ATTAAGGCCCTGTTTCATTAGAAATTTGTCCTTCCCTGCTTCAATGTTGTAAAGAT 47

Qy      425 TGTTCGTGTAAATATGTCCTTTATAATAACAGTAAAGCTG 467
      46 TGTTCGTGTAAATATGTCCTTTATAATAACAGTAAAGCTG 4

RESULT 16
LOCUS   R60026/c
DEFINITION YH12C03.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:43014 3' similar to contains MSRI repetitive element ;, mRNA
sequence.
ACCESSION R60026
VERSION   R60026
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 492)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
TITLE     The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 355
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: SP6
High quality sequence stop: 363.
FEATURES             source
     1..492
     /organism="Homo sapiens"
     /mol_type="mrna"
     /db_xref="GDB:415555"
     /db_xref="taxon:9606"
     /clone="IMAGE:43014"
     /sex="female"
     /dev_stage="73 days post natal"
     /lab_host="DH10B (ampicillin resistant)"
     /clone_lib="Soares infant brain INIB"
     /note="Organ: whole brain; Vector: Lafmid BA; Site 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AAGTGAAGATTCGGCGCCGAGGAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN
Query Match      84.5%; Score 394.6; DB 14; Length 492;
Best Local Similarity 94.9%; Pred. No. 1.7e-92;
Matches 447; Conservative 0; Mismatches 20; Indels 4; Gaps 4;

Qy      1 CAATGTTGCTATCCACCTCCCGCCAGCCCTTACCTATG-CTGCTGTAAGCTGCT 59
      482 CCAATTTGCTTAACCACTCCCGCCAGCCCTTACCTAATGCTGCTCTAAAGCTGCT 423

```

QY 60 GCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGG-ACTGGTGGTG-CCAG 117

Db 422 GCTGCTGCTGCTGCTTAAGGCTCATGCTTGGAGTGGGAACTGGTGGTGCCAG 363

QY 118 AAAGTCTCTTCTG-CCACTGACGCGCCCATCAGGGAATGGGCCCTTCCTTCCCCCTTCCTT 176

Db 362 AAAGTCTCTTCTGCCCCTGACTGACGCCCATCAGGGAATGGGCCCTTCCTTCCCCCTTCCTT 303

QY 177 TCTGTGCTCTGCTGCTCATGGGCTGCATGACCTGCGAGCAACCCAGCCCGCTGGGGA 236

Db 302 TCTGTGCTCTGCTGCTCATGGGCTGCATGACCTGCGAGCAACCCAGCCCGCTGGGGA 243

QY 237 AGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTG 296

Db 242 AGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTG 183

QY 297 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATACATAATTTAT 356

Db 182 GTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATACATAATTTAT 123

QY 357 ACAAGATATTAAGGCCCTGTCATTAGAAATGTTCCCTTCCCTGCTGTTCAATGTTT 416

Db 122 ACAAGATATTAAGGCCCTGTCATTAGAAATGTTCCCTTCCCTGCTGTTCAATGTTT 63

QY 417 GTAAGATGTTCTGTGTAATATGTTCTTATATAAACAGTAAAAAGCTG 467

Db 62 GTAAGATGTTCTGTGTAATATGTTCTTATATAAACAGTTANAANTG 12

RESULT 17

AUI56068/c

LOCUS AUI56068 PLACE1 Homo sapiens cdna clone PLACE1002140 3', mRNA

DEFINITION sequence.

ACCESSION AUI56068

VERSION AUI56068.1 GI:11017589

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 403)

AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.).

TITLE HRI human cDNA project (Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Ozawa,M., Nakamura,Y., Nagai,T., Sugano,S., Isogai,T.)

JOURNAL Unpublished (2000)

COMMENT Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

source

1..403

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="PLACE1002140"

/tissue\_type="placenta"

/clone\_lib="PLACE1"

/note="Vector: pME18SFL3"

ORIGIN

Query Match 83.8%; Score 391.2; DB 9; Length 403;

Best Local Similarity		97.5%	Pred. No. 1.2e-91;	
Matches	393;	Conservative	0;	Mismatches 10; Indels 0; Gaps 0;
QY	65	TGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAAGTCT	124	
Db	403	TGCTGNTGCTGCTTTAAAGGCTCATGNTTGGAGTGGGANTGTTGGTGGCCAGAAAGTCT	344	
QY	125	CTTCTGCACCTGACGCGCCCATCAGGATGAGGATGAGGCTTCTTCCCTCCCTTCTTCTTGCTGTC	184	
Db	343	TTTGTGGCATGACGCGCCCATCAGGATGAGGCTTCTTCCCTCCCTTCTTCTTGCTGTC	284	
QY	185	TCCTGCTCATGCGGCTGCGCATGACCTGCAGCCAAAGCCAGCCGCGCTGGGGAAGGGGAGA	244	
Db	283	TCCTGCTCATGCGGCTGCGCATGACCTGCAGCCAAAGCCAGCCGCGCTGGGGAAGGGGAGA	224	
QY	245	AAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGATGAGGCTGGGCT	304	
Db	223	AAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGATGAGGCTGGGCT	164	
QY	305	AGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATATACTAATTTATACAAAGAT	364	
Db	163	AGGGGGCTGCTTATTTAAAGTGGTGTGTTATGATCTTATATACTAATTTATACAAAGAT	104	
QY	365	ATTAAGCCCTGTTCAATTAAGAAATTTGTTCCCTTCCCTGTTCAATGTTGTAAGAT	424	
Db	103	ATTAAGCCCTGTTCAATTAAGAAATTTGTTCCCTTCCCTGTTCAATGTTGTAAGAT	44	
QY	425	TGTTCTGTGTAATATGCTCTTTATAATAAAGCTG 467		
Db	43	TGTTCTGTGTAATATGCTCTTTATAATAAAGCTG 1		
<p>RESULT 18</p> <p>T27131/c</p> <p>LOCUS</p> <p>DEFINITION NIBT311E03 Infant brain, LML array of Dr. M. Soares IN1B Homo sapiens cDNA clone LLAB311E03 3'end, mRNA sequence.</p> <p>ACCESSION T27131</p> <p>VERSION T27131.1</p> <p>KEYWORDS GI:774173</p> <p>SOURCE Homo sapiens (human)</p> <p>ORGANISM</p> <p>REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>AUTHORS Sikela,J.M.</p> <p>TITLE Human cDNAs from infant brain</p> <p>JOURNAL Unpublished (1993)</p> <p>COMMENT Contact: Sikela JM Department of Pharmacology University of Colorado Health Sciences Center Box C236, 4200 E. 9th Ave, Denver CO 80262-0236 Tel: 3032708637 Fax: 3032707097 Email: nikki@ally.uchsc.edu Seq primer: -21M13 Universal.</p> <p>FEATURES</p> <p>source</p> <p>1. .414</p> <p>Location/Qualifiers</p> <p>/organism="Homo sapiens"</p> <p>/mol_type="mRNA"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="LLAB311E03"</p> <p>/clone_lib="Infant brain, LML array of Dr. M. Soares IN1B"</p> <p>/notes="Vector: lafmid BA; Site 1: HindIII; Site 2: NotI; Normalized infant brain cDNA library made by Dr. M. Soares (Columbia University), oligo-dT primed and directionally cloned between HindIII (5') and NotI (3') sites"</p> <p>ORIGIN</p> <p>Query Match 83.4%; Score 389.6; DB 14; Length 414;</p> <p>Best Local Similarity 97.1%; Pred. No. 3.2e-91;</p> <p>Matches 403; Conservative 0; Mismatches 11; Indels 1; Gaps 1;</p>				



```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1022YC23"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Query Match      81.5%; Score 380.6; DB 9; Length 977;
Best Local Similarity 95.5%; Pred. No. 9.7e-89;
Matches 420; Conservative 3; Mismatches 13; Indels 4; Gaps 3;

QY 1 CAATGTTGCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGCTA--ACGCTGC 58
Db 448 CAATGTTGCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTTAACGCTGCTGC 389
QY 59 TGCCTGCTGCTGCTGCTTAAGAGCTCATGCTTGAAGTGGGACTGCTGCTGCCAGA 118
Db 388 TGCCTGCTGCTGCTGCTTAAGAGCTCATGCTTGAAGTGGGACTGCTGCTGCCAGA 329
QY 119 AAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178
Db 328 AAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 269
QY 179 TGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 238
Db 268 TGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 209
QY 239 GCGAGAAAGTGGGGATGGCTAGAAAGCTGAGATAGGAGTGGGAGTGGGAGTGGG 298
Db 208 GCGAGAAAGTGGGGATGGCTAGAAAGCTGAGATAGGAGTGGGAGTGGGAGTGGG 149
QY 299 TGGGCTAGGGGGGCTGCTGCTTATTAAGTGGT--TGTTTATGATCTTATCTAATTTATA 357
Db 148 TGGGCTAGGGGGGCTGCTGCTTATTAAGTGGT--TGTTTATGATCTTATCTAATTTATA 89
QY 358 CAAAGATATTAAAGCCCTGTTTAAAGAAATGTTT--CCTCCCTGTTGTTCAATGTTT 416
Db 88 CAAAGATATTAAAGCCCTGTTTAAAGAAATGTTT--CCTCCCTGTTGTTCAATGTTT 29
QY 417 GTAAAGATTGTTCTGTGTAA 436
Db 28 GTAAAGATTGTTCTGTGTAA 9

```

```

RESULT 21
AW594286/c 386 bp mRNA linear EST 22-MAR-2000
LOCUS hg57h03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2949749 3',
DEFINITION mRNA sequence.
ACCESSION AW594286
VERSION AW594286.1 GI:7281544
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-i@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

```

```

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LENL at:
image.llnl.gov/image/html/iresources.shtml
Seq primer: -40UP from Gibco.

```

## FEATURES

```

Location/Qualifiers
1..386
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2949749"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP GC6"
/vector="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonids
1257096-1258531, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

```

## ORIGIN

```

Query Match      81.2%; Score 379.4; DB 10; Length 386;
Best Local Similarity 99.7%; Pred. No. 1.5e-88;
Matches 380; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 87 ATGCTTGGAGTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 146
Db 386 ATGCTTGGAGTGGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 327
QY 147 CAGGAGATTGGGCTTCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 206
Db 326 CAGGAGATTGGGCTTCTTCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 267
QY 207 GACCTGCACCAAGCCCGCCGCGGAGGAGAGTGGGGATGGGCTAAGAAAG 266
Db 266 GACCTGCACCAAGCCCGCCGCGGAGGAGAGTGGGGATGGGCTAAGAAAG 207
QY 267 CTGGGAGATAGGGAACAGAAAGAGGTAGTGGCTAGGGGGGCTGCTTATTAAAG 326
Db 206 CTGGGAGATAGGGAACAGAAAGAGGTAGTGGCTAGGGGGGCTGCTTATTAAAG 147
QY 327 TGGTGTGTTATGATCTTATCTAATTTATAAAGATATTAAAGCCCTGTTCAATTAAGA 386
Db 146 TGGTGTGTTATGATCTTATCTAATTTATAAAGATATTAAAGCCCTGTTCAATTAAGA 87
QY 387 AATTGTTCCCTTCCCTGTTCAATGTTTGTAAAGATTGTTCTGTGTAATATGCTTT 446
Db 86 AATTGTTCCCTTCCCTGTTCAATGTTTGTAAAGATTGTTCTGTGTAATATGCTTT 27
QY 447 ATAATAAACAGTTAAAGCTG 467
Db 26 ATAATAAACAGTTAAAGCTG 6

```

## RESULT 22

```

BQ024597/c 356 bp mRNA linear EST 27-MAR-2002
LOCUS UI-1-BB1p-aus-e-04-0-UI-s1 NCI CGAP Pl6 Homo sapiens cDNA clone
DEFINITION UI-1-BB1p-aus-e-04-0-UI 3', mRNA sequence.
ACCESSION BQ024597
VERSION BQ024597.1 GI:19759876
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Dr. Steven Brown  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
source  
1. .356  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UT-1-BB1p-aus-e-04-0-UI"  
/tissue\_type="Placenta"  
/dev\_stage="Full Term"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP Pl6"  
/note="Organ: Placenta; Vector: p77T3-Pac (Pharmacia) with  
a modified polylinker; Site 1: EcoR I; Site 2: Not I;  
NCI CGAP Pl6 is a subtracted cDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into p77T3-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are GA, AGGAA. For  
additional information, contact: Bento Soares,  
bento-soares@uiowa.edu  
TAG TISSUE=placenta human full term  
TAG\_LIB=UT-1-BB1p  
TAG\_SEQ=AGGAA"

ORIGIN

Query Match 72.4%; Score 338; DB 12; Length 356;  
Best Local Similarity 100.0%; Pred. No. 1e-77;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GCCACTGACGCCCATCAGGAGTGGCCCTCTTCCCTTCTCTTCTGTCTCTG 189  
DB 356 GCCACTGACGCCCATCAGGAGTGGCCCTCTTCCCTTCTCTTCTGTCTCTG 297

QY 190 CCTCATCGGCTGCATGACCTGCAGCCCAAGCCCGCTGGGAGGGAAGAGTG 249  
DB 296 CTTATCGGCTGCATGACCTGCAGCCCAAGCCCGCTGGGAGGGAAGAGTG 237

QY 250 GGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTGTAGTGGGTAGGG 309  
DB 236 GGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTGTAGTGGGTAGGG 177

QY 310 GGCTGCCTTATTTAAAGTGGTGTATGATCTTATCTAATTTATACAAAGATATTA 369  
DB 176 GGCTGCCTTATTTAAAGTGGTGTATGATCTTATCTAATTTATACAAAGATATTA 117

QY 370 GGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTGTGTGTCAATGTTTAAAGATTGTC 429  
DB 116 GGCCCTGTTTCATTAAGAAATGTTCCCTTCCCTGTGTGTCAATGTTTAAAGATTGTC 57

QY 430 TGTGTAATATGCTTTTAAATTAACAGTAAAGCTG 467  
DB 56 TGTGTAATATGCTTTTAAATTAACAGTAAAGCTG 19

RESULT 23  
AI886944/c  
LOCUS  
DEFINITION w194d03.x1 NCI CGAP\_Brn25 Homo sapiens cDNA clone IMAGE:2432549 3',

mRNA sequence.  
AI886944 GI:5592108  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 612)  
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGP), Tumor Gene Index  
Unpublished (1998)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 517 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 312.  
Location/Qualifiers  
1. .612  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2432549"  
/tissue\_type="anaplastic oligodendroglioma"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Brn25"  
/note="Organ: brain; Vector: p77T3D-Pac (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [S'  
TGTACCAATCTGAAGTGGAGCGCCGATAGTGTCTTTTCTTTTCTTTTCTTTTCTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified p77T3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 65.0%; Score 303.6; DB 9; Length 612;  
Best Local Similarity 93.1%; Pred. No. 1.3e-68;  
Matches 363; Conservative 0; Mismatches 19; Indels 8; Gaps 4;

QY 78 TAAAGGCTCATGTCTGAGTGGGAGCTGTCGTCGCCAGAAAGTCTCTCTGCACTGA 137  
DB 384 TAAAGGCTCATGTCTT---GAGTGGGACTGTTCGTCGCCAGAGGTCTCTCT--CTGCACTGA 330

QY 138 CGCCCCCATCAGGATTTGGCCCTTCTTCCCTTCTCTTCTGTCTCTGCTCATCG 197  
DB 329 CCCCCCATCAGGAG-TTGGGCTTCTTCTTCCCTTCTCTTCTCTCTCTCATCG 273

QY 198 GCCTGCATCATCTGCAGCCCAAGCCCGCTGGGGAAGGGGAGAAAGTGGGGATGG 257  
DB 272 GCCTGCATCATCTGCAGCCCAAGCCCGCTGGGGAAGGGGAGAAAGTGGGGATGG 213

QY 258 CTAGAAAGCTGGAGATAGGGAACAGAGAGGGTGTAGTGGGTAGGGGGGCTGCCT 317  
DB 212 CTAGAAAGCTGGAGATAGGGAACAGAGAGGGTGTAGTGGGTAGGGGGGCTGCCT 153

QY 318 TATTTAAAGTGGTGTATGATCTTATCTAATTTATACAAAGATATTAAAGCCCTGT 377  
DB 152 TATTTAAAGTGGTGTATGATCTTATCTAATTTATACAAAGATATTAAAGCCCTGT 93

QY 378 TCATTAAGAAATGTTCCCTTCCCTTCTCTCAATGTTTCTAAAGATTGTTCTGTGTAAA 437



```

QY 377 TCCATTAAGAAATGTCCTCCCTCCCTGGTGTCAATGTTGTAAGAGATTGTTCTGTGTA 436
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 91 TCCATTAAGAAATGTCCTCCCTCCCTGGTGTCAATGTTGTAAGAGATTGTTCTGTGTA 32
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 437 ATATGCTCTTTATAATAACAGATTAAAGCTG 467
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 31 ATATGCTCTTTATAATAACAGATTAAAGCTG 1
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 26
AA304416 334 bp mRNA linear EST 18-APR-1997
LOCUS EST17255 Aorta endothelial cells, TNF alpha-treated Homo sapiens
DEFINITION cDNA 5' end, mRNA sequence.
ACCESSION AA304416
VERSION AA304416.1 GI:1956820
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 334)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fink,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.B., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudke,D.M., Shikley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarik,D.P., Gao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen.H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
7566098
Other ESTs: THC174412
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@igr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1..334
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="ATCC (inhost):116104"
/db_xref="taxon:9606"
/cell_type="endothelial cell"
/dev_stage="adult"
/clone_lib="Aorta endothelial cells, TNF alpha-treated"
/note="Organ: aorta; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

ORIGIN
Query Match 56.1%; Score 261.8; DB 9; Length 334;
Best Local Similarity 96.0%; Pred. No. 9.7e-58;
Matches 290; Conservative 0; Mismatches 8; Indels 4; Gaps 2;

```

```

QY 1 CAATGTTTCCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGTAACGCTGCTG 60
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 16 CAATGTTTCCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGTAACGCTGCTG 75
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 61 CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAA 120
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 76 CTGCTGCTGCTGCTGCTTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAA 135
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 121 GTCTCTTCGCACTGACGCGCCCATCAGGATTTGGGCTTCTTTCCGCCCTTCCTTTCTG 180
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 136 GTCTCTTCGCACTGACGCGCCCATCAGGATTTGGGCTTCTTTCCGCCCTTCCTTTCTG 195
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 181 TGTCTCTTCGCTCATCGGCTGCGCATGACCTGACGCAAGCCAGCCCGGTGGGGAAGGG 240
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 196 TGTCTCTTCGCTCATCGGCTGCGCATGACCTGACGCAAGCCAGCCCGGTGGGGAAGGG 255
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 241 GAGAAAGT-GGGGGATGGCTAAGAAAGCT---GGGAGATAGGGAAACAGAGAGGGTAGTG 296
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 256 GAGAAAGTGGGGGGATGGCTAAGAAAGCTTGGAGATTAGGGGACCAGAGAGGGTAGTT 315
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

QY 297 GG 298
Db ||
QY 316 GG 317
Db ||

RESULT 27
BF939932/c 300 bp mRNA linear EST 22-JAN-2001
LOCUS nacc64d01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3439129
DEFINITION 3', mRNA sequence.
ACCESSION BF939932
VERSION BF939932.1 GI:12357252
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 300)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 292.
Location/Qualifiers
1..300
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3439129"
/cell_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP Brn23"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTCACCACTGTAAGTGGGCGGCATATCTTTTCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.

```







/Clone\_lib="Canis Total Brain cDNAs"  
/note=vector: Lambda Zap II; The library was provided by  
Greg Hannon and Lee Santos (Cold Spring Harbor  
Laboratory). This library is oligo(dT) primed using  
stratagene zap cDNA synthesis kit. It was made from dog  
whole brain cells. please contact Greg Hannon  
(hannon@cshl.org) with any library related inquiries."

## ORIGIN

Query Match	37.6%;	Score 175.6;	DB 13;	Length 574;
Best local Similarity	72.3%;	Pred. No. 4.6e-35;		
Matches 298;	Conservative 0;	Mismatches 79;	Indels 35;	Gaps 4;

  

QY	63	GCTGCTCTGTCGTCTAAAGGCTCATGCTTGGAGTGGGAGCTGTCGTCGCCCAAGAAAGT	122
DB	150	GCACCTGCTTCCCCCAACACCCCGTGTCTGGATGTGGGACTGGCCAGGGTCCAGAAAGC	209
QY	123	CTCTTTCGCACATGACGCCGCCATCAGGAGATGGGCCTTCTTTCCCG-CTTCTCTTCTCGT	181
DB	210	CTCTTCTGTCACTGATGCCGCCCAACAGGAGCTGGGCCCTTCTTCCCTCTCCTCTCTGCG	269
QY	182	GTCTCTTCGCTCATCGGCGCTGCCATGACCTGCACCAAGCCACCCCGCTGGGGAGAGGGG	241
DB	270	GTCTCTTCGCTCATCAGC-----TGGCAGGCCCTCTGGGGAAAGGG	310
QY	242	AGAAAGTGGGGGATGGCTAAAGAAAGCTGGGAGATAGGGAAACAGAAAGGGGTAGTGGGTGG	301
DB	311	GGTAAAGCTGTAAAGTCCACAGACACACAGGCGCAGAAAGGGGCGAGTGGGGGGAG-----	363
QY	302	GCTAGGGGGGCTGCCCTATTTTAAAGTGGTTGTTTATGATCTCTATACTAATTTATACAAA	361
DB	364	--CTGGGGGCGCTGCTTATTTTAAAGTGGTTGTGATGATCTTATACTAATTTATACAAA	421
QY	362	GATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTTCCC-----CTGTGTTTCAAGTGT	415
DB	422	GATATTAAAGGCCCTTTTCATTACAGAAATGTGCCCTTCCCAATAATGTGTTCACTGTGTT	481
QY	416	TGTTAAAGATTTGTTCTGTGTTAAATATGTCCTTTATAATAATAACAGTTAAAGCTG	467
DB	482	TGTTAAAGATTTGTTCTGTATAAATATGTCCTTTATAATAATAAGAGTTAAAGCTG	533

RESULT 32	
BE101709/c	
LOCUS	
DEFINITION	linear EST 13-JUN-2000
	399 bp mRNA
	UI-R-B01-aps-c-11-0-UI.s1 UI-R-B01 Rattus norvegicus cDNA clone
	UI-R-B01-aps-c-11-0-UI 3', mRNA sequence.
ACCESSION	BE101709
VERSION	BE101709.1 GI:8493807
KEYWORDS	EST.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.

REFERENCE	1 (bases 1 to 399)
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
PMID	889548
COMMENT	Contact: Soares. MB

clones will be available through Research Genetics ([www.resgen.com](http://www.resgen.com))  
Seq primer: M13 Forward  
POLYA=Yes.

## ORIGIN

Query Match	31.7%	Score 148.2	DB 10	Length 399
Best Local Similarity	72.1%	Pred. No. 6.3e-28		
Matches 253	Conservative 0	Mismatches 83	Indels 15	Gaps 4
QY	121	GTCTCTCTGCGCACTGACGCCCCCATCAGGGATGGGCGTCTCTTCCCCCT-TCCTTTCT	179	
Db	357	GCCTCTTCTACTGCTGATCCCTGCGCAGGACGAGCCCTCTCTGTCTGTGTGTTTT	298	
QY	180	GTCTCTCTGCGCTCATCGGCGTCCCATGACTGCAGCCAGCCGACGCCGTGGGAAGG	239	
Db	297	GTACCTCTTGCCCTATCAGCGTCCATGGCCATCCACAGGGNAGG-----GG	249	
QY	240	GGAGAAAGTGGGGATGGCTTAAAGAGCTGGGAGATAGGGAAACAGAGAGGTAGTGGGT	299	
Db	248	AGGGAAGGAGGAGTGTGGCTGAGAAAGAGAGAGATAGAAACAGAGAGGGGAGTCAAT	189	
QY	300	GGCGTAGGGGGCGTCTTATTTAAAGTGGTGTGTTTATGATCTCTTACTAAATTTATACA	359	
Db	188	GGACCCAGTGGGCTGTCTTATTTAAAGTGGTGTGTATGATCTCTTACTAAATTTATATA	129	
QY	360	AAGATATTAAGGCCCTGTTCAATT-AAAGAAATGTTCCTTCCCTCCCTGTGTTTCA--ATGTTT	416	
Db	128	GAGATATTAAAGGCCCTTGAGTTAAAGAACTGTCTCTCATCCGCTGTGTTCACTATGTTT	69	
QY	417	GTAAGATGTGTCGTGTAAATATGTCTTTTATAATAAACAGTTAAAGCTG	467	
Db	68	GTAAAAATTTGTTCCATGTAAATATGTCTTTTATAATAAAGAGTTAAAGGCTG	18	

RESULT 33	ACCESSION	ORGANISM
CF109372	VERSION	
LOCUS	KEYWORDS	
DEFINITION	SOURCE	

ORGANISM

Rattus nor. vegatus  
Eukaryota; Metazoa;  
Mammalia; Eutheria;  
Rattus.  
1 (bases 1 to 799)  
Shultz, M.A., Zhang, L., Gu, Y.-Z., Baker, G.L., Fannuchi, M.V., Padua, A.M., Gurske, W.A., Morin, D., Penn, S.G., Jovanovich, S.B., Plapper, C.G. and Buckpitt, A.R.

REFERENCE

AUTHORS





State Univ., from 2 l; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Patima Bonaldo.

## ORIGIN

```

Query Match      27.8%; Score 129.8; DB 13; Length 520;
Best Local Similarity 69.0%; Pred. No. 4.6e-23;
Matches 245; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

Qy 121 GTCTCTTCTGCACTGACGCCGCCCATCAGGATTTGGGCTTCTTCCCCCT-TCCTTTCT 179
Db 123 GCCTCTTCTACTGCTGGTACCCCTACCAGGACCGGCTTCTCCGTTTGTGCTTCT 182
Qy 180 GTGTCTCTGCTCATCGCCCTGCTGACCTGACCTGAGCCAGCCAGCCCGCTGGGGAAGG 239
Db 183 GCACCTCTGCTTATCAGCTGCTGATGCCAGCCGCCACAGGAAAGGGGGGAATA 242
Qy 240 GGAGAAAGTGGGGATGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGT 299
Db 243 T-----TGGCTGAGAAAGAGGAGATAGA-AGCACAAGAGGGGAGTAAAC 287
Qy 300 GGGCTAGGGGGCTGCCCTATTAAAGTGGTGTATGATTTCTTATTAATAATTATACA 359
Db 288 GGGCCAGTGGGCTGCTTATTAAAGTGGTATGTGTATTTCTTATTAATAATTATATA 347
Qy 360 AAGATATTAAAGCCCTGTTTCAATTAAAGAAATGTTCCCTTCCC-----TGTGTCAAT 412
Db 348 GAGATATTAAAGCCCTTTGAGTAGAGAAAGCTGCCCTTCCCATAAGTGTGCTGCTAT 407
Qy 413 GTTTGTAAGATTGTTCTGTGTAATATATCTTTATTAATAACAGTTAAAGCTG 467
Db 408 GTTTGTAAGAAATGTTCCATGTAATATGTTTATTAATAAGAGTTAAAGTTG 462

```

```

RESULT 37
BI967057
LOCUS      521 bp      mRNA      linear      EST 12-MAR-2002
DEFINITION id27h04.y1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus
            musculus cDNA clone IMAGE:5664607 5', mRNA sequence.
ACCESSION  BI967057
VERSION     BI967057.1 GI:16341462
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE
AUTHORS    Melton,D., Brown,J., Kenty,G., Permuth,A., Lee,C., Kaestner,K.,
            Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
            Endocrine Pancreas Consortium
            Unpublished (2000)
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu

```

TITLE  
JOURNAL  
COMMENT

```

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
MGI:1950933 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
High quality sequence stop: 429.

```

FEATURES  
source

```

Location/Qualifiers
1. .521
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5664607"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH10B"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
N1-MMS1"
/notes="Vector: pSPORT1; Site.1: Not I; Site.2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
SuperScript plasmid library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an EcoT of 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library."

```

## ORIGIN

```

Query Match      27.8%; Score 129.8; DB 12; Length 521;
Best Local Similarity 69.0%; Pred. No. 4.6e-23;
Matches 245; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

Qy 121 GTCTCTTCTGCACTGACGCCGCCCATCAGGATTTGGGCTTCTTCCCCCT-TCCTTTCT 179
Db 148 GCCTCTTCTACTGCTGTAACCCCTACCAGGACCGGCTTCTCCGTTTGTGCTTCT 207
Qy 180 GTGTCTCTGCTCATCGGCTGCTGACCTGAGCCAGCCAGCCCGCTGGGGAAGG 239
Db 208 GCACCTCTGCTTATCAGCTGCTGCGCCAGCCGCCACAGGAAAGGGGGGAATA 267
Qy 240 GGAGAAAGTGGGGATGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGT 299
Db 268 T-----TGGCTGAGAAAGAGGAGATAGA-AGCACAAGAGGGGAGTAAAC 312
Qy 300 GGGCTAGGGGGCTGCCCTATTAAAGTGGTGTATGATTTCTTATTAATAATTATACA 359
Db 313 GGGCCAGTGGGCTGCTTATTAAAGTGGTATGTGTATTTCTTATTAATAATTATATA 372
Qy 360 AAGATATTAAAGCCCTGTTTCAATTAAAGAAATGTTCCCTTCCC-----TGTGTCAAT 412
Db 373 GAGATATTAAAGCCCTTTGAGTAGAGAAAGTGTCCCTTCCCATAAGTGTGCTGCTAT 432
Qy 413 GTTTGTAAGATTGTTCTGTGTAATATGTTTATTAATAACAGTTAAAGCTG 467
Db 433 GTTTGTAAGAAATGTTCCATGTAATATGTTTATTAATAAGAGTTAAAGTTG 487

```

```

RESULT 38
CF582778
LOCUS      966 bp      mRNA      linear      EST 24-SEP-2003
DEFINITION AGNCOURT_11363244 updated NIH_MGC_137 Mus musculus cDNA clone
            IMAGE:6432441 5', mRNA sequence.
ACCESSION  CF582778
VERSION     CF582778.1 GI:35196040
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```



LIGAND), full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AK042076  
GI:26334910  
HTC; CAP trapper  
Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

REFERENCE

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

REFERENCE

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2604)

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),

RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,

Kanagawa 230-0045, Japan (E-mail: genome-research@gs.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics

Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome  
Trust/MRC building Addenbrookes Hospital Cambridge) whose  
assistance we gratefully acknowledge.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1. 2604

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:A630056D08"

/db\_xref="MGI:2407093"

/db\_xref="taxon:10090"

/clone="A630056D08"

/tissue\_type="thymus"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="3 days neonate"

352.585

/notes="unnamed protein product; APELIN PRECURSOR (APJ

ENDOGENOUS LIGAND) (SWISSPROT|Q9R0R4, evidence: FASTY,

100%ID, 100%length, match=231)

putative"

/codon\_start=1

/protein\_id="BAC31156.1"

/db\_xref="GI:26334911"

/translation="MNLRLCVQALLLWSLTAVCVGLMPLPDGTGLEBGRMYLVK

PRTSRGPAGWQGRKRRKRRRLSHKGPMPF"

2587.2592

/notes="putative"

2604

/notes="putative"

polyA\_site

polyA\_site

ORIGIN

Query Match 27.6%; Score 128.8; DB 11; Length 2604;

Best Local Similarity 68.3%; Pred. No. 1.4e-22;

Matches 244; Conservative 0; Mismatches 87; Indels 23; Gaps 4;

121 GTCTCTTCTGCCACTGACGCCCTCCATCAGGAGTGGGCTTTCTTCCCTTCT 179

2266 GCCTCTTCTTCTGCTGCTACCTCCCTACCGAGCGGCTTCTTCTTCTTCT 2325

180 GTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 239

2326 GCACCTCTCTGCTTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2385

240 GGAGAAAGTGGGGTGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGCTAGTGGGT 299

2386 T-----TGGCTGAGAAAGAGAGAGATAGA-AGCACAAGAGGGGAGTAAAC 2430

300 GGCTAGGGGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 359

2431 GGCGCCAGTGGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2490

360 AAGATATTAAAGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 412

2491 GAGATATTAAAGGCTTGTAGTAGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2550

413 GTTTGTAAGATTTGTTCTGTGTAATATGCTTTTATAATAAAGCTTAAAGCT 466

2551 GTTTGTAAGATTTGTTCTGTGTAATATGCTTTTATAATAAAGCTTAAAGCT 2604

RESULT 41

BI251779

DEFINITION

mRNA sequence.

ACCESSION

BI251779

VERSION

BI251779.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

BT251779 602994285F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:5149940 5',

780 bp mRNA linear EST 17-JUL-2001





SOURCE ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 596)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 630 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 394.

FEATURES  
source  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:1268316"  
/db\_xref="taxon:9606"  
/clone="IMAGE:342941"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares fetal heart NBH119W"  
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGCGGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH119W."

ORIGIN  
Query Match 27.2%; Score 127; DB 14; Length 596;  
Best Local Similarity 100.0%; Pred. No. 2.6e-22;  
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 TCTTATCTAATTATACAAAGATATTAAAGCCCTGTTCATTAAAGAAATGTTCCCTTCC 400  
Db 1 TCTTATCTAATTATACAAAGATATTAAAGCCCTGTTCATTAAAGAAATGTTCCCTTCC 60

Qy 401 CCTGTGTCATGTTGTGAAGATGTTCTGTGTAATATGTCCTTATATAACAGTTA 460  
Db 61 CCTGTGTCATGTTGTGAAGATGTTCTGTGTAATATGTCCTTATATAACAGTTA 120

Qy 461 AAAGCTG 467  
Db 121 AAAGCTG 127

RESULT 44  
LOCUS W84234  
DEFINITION mf36f06.rl Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA clone IMAGE:407171 5', mRNA sequence.  
ACCESSION W84234  
VERSION W84234.1 GI:1541218  
KEYWORDS EST.

SOURCE ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 486)  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisels, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395365.  
Contact: Marra M/Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:250939  
Seq primer: -29M13 rev2 from Amersham  
High quality sequence stop: 484.

FEATURES  
source  
1..486  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:407171"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NBME13.5 14.5"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGCGGCATCTTTTCTTTTCTTTT 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN  
Query Match 27.1%; Score 126.6; DB 14; Length 486;  
Best Local Similarity 68.5%; Pred. No. 3.1e-22;  
Matches 243; Conservative 0; Mismatches 89; Indels 23; Gaps 4;

Qy 121 GTCTCTTCTGCCACTGACGCCCTCATCAGGATGCGCTTCTTCCCTTCTTCTTCT 179  
Db 121 GCCTCTTCTACTGCTGGTACCCCTTACCAGGACCGGCGCTTCTCCGTTTGTGTTCT 180

Qy 180 GTGTCTCTCTCCCTCATCGGCTTGCCTGACCTGACGACCAAGCCCGCCGCTGGGGAAGG 239  
Db 181 GCACCTCTCTCCCTTATCAGCTGCCATGGCCGCCACAGGGAAGGGGGGAATA 240

Qy 240 GGAGAAAGTGGGGATGCTAAGAAAGCTGGGAGATAGGGAACAGACAGAGGTAGTGGGT 299  
Db 241 T-----TGGCTGAGAAAGAGAGAGATAGA-AGCACAGAGAGGGAGTAAC 285

Qy 300 GGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATCTTATATAAATTATACA 359  
Db 286 GGGCCCCAGTGGCGTGTCTTATTTAAAGTGGTGTGTTATGTTATCTTATATAAATTATATA 345

Qy 360 AAGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTC-----TGTGTTCAAT 412

```

Db      346 GAGATATTAAGGCCCTTGGAGTTAGAGAAACTGTCCCTTCCCATAGTGTGTACGGCTAT 405
Qy      413 GTTTGTTAAAGATTCTTCTGTGTAATAATGCTTTTATATAAACAAGTTAAAGCTG 467
      |||||
Db      406 GTTTGTTAAAGATTCTTCCATGTAATAATATGCTTTATATAAAGAGTTAAAGATTG 460

RESULT 45
BB819942 449 bp mRNA linear EST 19-NOV-2001
DEFINITION BB819942 RIKEN full-length enriched, mammary gland RCB-0526
            Jyg-MC(A) cDNA Mus musculus cDNA clone G830003B15 3', mRNA
            sequence.
BB819942
ACCESSION BB819942.1 GI:16992571
VERSION BB819942.1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (Bases 1 to 449)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiracka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komno,H., Kouda,M.,
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nomasaki,R.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toyota,T., Watahiki,A., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
TITLE Unpublished (2001)
JOURNAL Contact: Yoshihide Hayashizaki
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shero-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1. 449
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G830003B15"
/tissue_type="mammary gland"
/cell_line="RCB-0526 Jyg-MC(A)"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0526 Jyg-MC(A) cDNA"

FEATURES
SOURCE
1. 449

ORIGIN
1. 342

```

```

Query Match 26.8%; Score 124.2; DB 10; Length 449;
Best Local Similarity 68.7%; Pred. No. 1.3e-21;
Matches 252; Conservative 0; Mismatches 93; Indels 22; Gaps 5;

Qy      106 GTCGGTGGCCAGAAAGTCTCTCTGCGCACATGACGCCCCCATCAGGGATTGGGCTTCTTT 165
      |||||
Db      95 GTTGGAGCTGGCTGGCGCTCTTCTACTGTGTGTACCCCTACAGGACCGGGCTTCTCC 154
Qy      166 CCCCTT-TCCTTTCTGTGTCTCTGCTCATCGGCTGCGCATGACCTGCGAGCCAAAGCCCA 224
      |||||
Db      155 GTCTTTGTCTGTTCTGCACTCTCTTCTTATCATCAGCCTGCCATG-----GCCAGGCCCC 207
Qy      225 GCCCGTGGGGAAGGGAGAAAGTGGGGGATGCTAGAAAGCTGGGAGATAGGGAACAG 284
      |||||
Db      208 ACAGGGAAGGGAGGGGGGAATAT-----TGCTGAGAAAGAGGAGATAGA-AGCAC 260
Qy      285 AAGAGGGTAGTGGTGGCTAGGGGGGCTGCCTTATTTAAAGTGGTTGTTATGATTCCT 344
      |||||
Db      261 AAGAGGGGAGTAAACGGGCCAGTGGGCTGCTTATTTAAAGTGGTTATGTTATTCCT 320
Qy      345 ATACTAATTTATACAAAGATATTAAAGCCCTGTTCATTAGAATAATGTTCCCTTCCCC-- 402
      |||||
Db      321 ATACTAATTTATATAGAGATATTAAAGCCCTTTGAGTTAGAGAACTGTCCCTTCCCAT 380
Qy      403 -----TGTGTTCAATGTTTCTAAAGATTGTTCTGTGTAATATGCTTTTATAATAACAG 457
      |||||
Db      381 AAGTGTGTTGCGTATGTTTTTAAANAATCTTCCATGTAATAATGCTTTTATAATAAGAG 440
Qy      458 TTAAAG 464
      |||||
Db      441 TTAAAG 447

RESULT 46
BF600467 342 bp mRNA linear EST 25-APR-2001
LOCUS 265082 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BF600467
ACCESSION BF600467
VERSION BF600467.1 GI:11697624
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (Bases 1 to 342)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Fertez,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
JOURNAL MEDLINE
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTCCCACTCAGCAGC
Plate: 37 row: P column: 11
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. 342

```

[illegible]

/dev> save="20 week-post conception fetus"  
/lab host="DH10B (ampicillin resistant)"  
/cclone lib="Soares fetal liver spleen INFLS S1"  
/clone="Organ: Liver and Spleen; Vector: pT73D (Pharmacla)  
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;  
This is a subcloned version of the original Soares fetal  
liver spleen INFLS library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5'  
AAGTGGAGAAATTAATGAAGATCTTTTCTTTTCTTTT 3'],  
(Pharmacla), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

Query Match	25.3 %;	Score 118;	DB 14;	Length 616;
Best Local Similarity	100.0 %;	Fred. No. 6e-20;		
Matches 118;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

  

QY	350	AATTATATACAAAGATATTAAAGCCCTGTGTTCAATAAGAAATTGTTCCCTTCCTCGTGTTC	409
DB	1	AATTTTATACAAAGATATTAAAGCCCTGTGTTCAATAAGAAATTGTTCCCTTCCTCGTGTTC	60
QY	410	AATGTTTGTAAAGATTGTTCTCTGTGTAATAATGCTTTATAATAACAGTTAAAGCTG	467
DB	61	AATGTTTGTAAAGATTGTTCTCTGTGTAATAATGCTTTATAATAACAGTTAAAGCTG	118

and -mismatch 12 options.

PCR Primers  
 FORWARD: AGGAACAGCTATGACACG  
 BACKWARD: GTTTCCTCAGTCACGACG  
 Plate: 105 row: I column: 14  
 Seq primer: ATTAGGTGACACTATAG.

FEATURES  
 source  
 1. .480  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /tissue type="pooled"  
 /lab host="DH10B"  
 /clone lib="BARC 5BOV"  
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
 Library made from pooled mRNA isolated from mammary  
 tissues at eight physiological, developmental, and disease  
 states."

## ORIGIN

Query Match 25.2%; Score 117.8; DB 10; Length 480;  
 Best Local Similarity 69.6%; Pred. No. 6.2e-20;  
 Matches 218; Conservative 0; Mismatches 62; Indels 33; Gaps 3;  
 QY 89 GCTTGGAGTGGGACTGGTGGTCCAGAAAGTCTCTTCTGCCACTTGACGCCGCCCATCA 148  
 Db 201 GCCCGAGTGGGGCTGACCAAGTGCAGAAAGCTCTCTGCCATGATGCCGCCACCA 260  
 QY 149 GGAATGGGCTCTTTCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 208  
 Db 261 GGGACTGGGCGTTTTCCT 310  
 QY 209 CCTGCACCAAGCCAGCCCGTGGGAGGGGAGAGTGGGGATGGCTAGAAAGCT 268  
 Db 311 -----TTGCCAGCCCTGTGGGAA--CGAGATGGAAGGGATGGCTGGGAAGTC 359  
 QY 269 GGGAGATAGGGAACAGAAAGGAGTGGGTGGGCTAGGGGGCTGCTCTATTATAAGTG 328  
 Db 360 TGG-----GGACAGAGAGTGGGGATGCTGAGGGCTGTCTATTATAAGTG 407  
 QY 329 GTTGTATTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTAT 388  
 Db 408 GTTGTGTGATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTAT 467  
 QY 389 TTCTTCCTTCCC 401  
 Db 468 GCCTTCCCTTCC 480

## RESULT 49

BI455596 921 bp mRNA linear EST 21-AUG-2001  
 LOCUS 60317484F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:5257860 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI455596  
 VERSION BI455596.1 GI:15246252  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov  
 Plate: LLAM11651 row: k column: 13  
 High quality sequence stop: 655.

FEATURES  
 source

1. .921  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5257860"  
 /tissue type="tumor, gross tissue"  
 /dev stage="7 months"  
 /lab host="DH10B"  
 /clone lib="NCI\_CGAP Mam5"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

## ORIGIN

Query Match 24.8%; Score 115.8; DB 12; Length 921;  
 Best Local Similarity 69.9%; Pred. No. 2.6e-19;  
 Matches 248; Conservative 0; Mismatches 87; Indels 20; Gaps 6;  
 QY 121 GTCTCTTCTGCCACTGACGCCGCCCATCAGGATGGGCTTCTTCCCTTCTTCTCTCTCTCT 180  
 Db 113 GCCTCTTCTACTGCTGTACCCCTACCGGAGCCGGGCTTC-----TCTGTCTT 163  
 QY 181 TGCTCTCTCTCATCGGCTGCCATGACTGCAGCCAAAGCCAGCCCGTGGGGAAGGG 240  
 Db 164 TGCTGTCTTCTGCA--CTCTCTGCTTATCAGCTGCGCTGCCAGCCACAGGGAAGG 222  
 QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGAAACAGAGAGGATGGGTTG 300  
 Db 223 GAG--GGGGGAATATTGGCTGAGAAAGAGGAGAGATAGA--AGCACAGAGAGGAGTAAACG 280  
 QY 301 GCTAGGGGGCTGCTTATTAAAGTGGTGTGTTATGATCTTATCTTATCTTATCTTATCTTAT 360  
 Db 281 GGCCCAAGTGGGCTGCTTATTAAAGTGGTGTGTTATGATCTTATCTTATCTTATCTTATCT 340  
 QY 361 AGATATTAAAGGCTGTTTCATTAAAGAAATGTTCCCTTCCCTCCCT-----TGTTTCAATG 413  
 Db 341 AGATATTAAAGGCTTGGATTAGAGAACTGCTCCCTTCCCAATGATGTTGCTGATG 400  
 QY 414 TTGTGAA--AGATTGTTCTGTCTGTAATATATCTTATTAATAAAGTAAAGCTG 467  
 Db 401 TTGTGAAATTTGTTCCATGTAATATCTTATTAATAAAGAGTAAAGTTG 455

## RESULT 50

AV365977 231 bp mRNA linear EST 14-NOV-1999  
 LOCUS AV365977 RIKEN full-length enriched, 16 days embryo lung Mus  
 DEFINITION musculus cDNA clone 8430414M03 3', mRNA sequence.  
 ACCESSION AV365977  
 VERSION AV365977.1 GI:6413624  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 231)  
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,  
 Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,  
 Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, T.,  
 Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Owa, C., Orawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K.,  
 Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Suganara, Y.,  
 Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimaga, N.,  
 Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A.,  
 Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

# TITLE JOURNAL COMMENT

RIKEN Mouse ESTs (Konno,H., et al. 1999)  
Unpublished (1999)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki.N., Izawa.M., Watahiki.M., Ozawa.K., Tanaka.T., Yoneda.Y.,  
Matsuura.S., Carninci.P., Muramatsu.M., Okazaki.Y. and  
Hayashizaki.Y.

Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh.M., Kitsuai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,  
Tomaru.Y., Carninci.P., Shibata.Y., Ozawa.Y., Muramatsu.M.,  
Okazaki.Y. and Hayashizaki.Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci.P. and Hayashizaki.Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

## FEATURES source

Location/Qualifiers  
1..231  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="8430414M03"  
/sex="mixed"  
/tissue\_type="lung"  
/dev\_stage="16 days embryo"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 16 days embryo  
lung"  
/note="Site 1: SalI; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 185.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATCTCGAGTTAATTAATATCCCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified phuescript KS(+) after bulk excision  
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:  
BamHI"

## ORIGIN

Query Match 23.0%; Score 107.2; DB 9; Length 231;  
Best Local Similarity 74.5%; Pred. No. 2.9e-17;  
Matches 164; Conservative 0; Mismatches 48; Indels 8; Gaps 2;  
QY 255 TGGCTAAGAAAGCTGGAGATAGGAAACAGAGAGGAGTAGTGGGTGGCTAGGGGGGCTG 314  
DB 13 TGTCTAAGAAAGAGAGAGATA-GAAGCACAGAGGGGAGTAACGGGCCCGAGTGGGCTG 71  
QY 315 CCTATTATTAAGTGGTCTTATGATCTTATCTAATTAATTAACAAGATATTAAAGGCC 374  
DB 72 TCTTATTAAAGTGGTATGTGTTCTTATCTAATTAATTAATAGATATTAAAGCCC 131  
QY 375 TGTTCATTAAAGAAATGTTCCCTCCCC-----TGTGTTCAATGTTGTAAAGATTGT 427

Db 132 TTTGAGTTAGAGAAACTGTCCCTTCCCATAGTGTTCGCTATGTTTGTAAAAATTCT 191  
QY 428 TCTGTGTAATATGTCCTTTTATAAACAAGTTAAAGCTG 467  
DB 192 TCCATGTAATAATTTTTTTTATAATAAAGAGTTAAAGTTG 231

Search completed: May 25, 2004, 17:18.12  
Job time : 2869 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 17:18:40 ; Search time 2864 Seconds  
(without alignments)  
4869.287 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caaagtgtgcctaccct.....taataaacagttaaaagctg 467

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 10

Total number of hits satisfying chosen parameters: 13639910

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_eston:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_nam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	466	99.8	728	12	BM975075
C 2	447	95.7	450	9	AI149981
C 3	414	88.7	449	9	AW015212
C 4	414	88.7	517	14	CA444588

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C 5	414	88.7	568	12	BQ006616
C 6	384	82.2	555	10	BE855577
C 7	373	79.9	889	13	EX434223
C 8	370	79.2	431	9	AA099445
C 9	363	77.7	687	14	CA438256
C 10	347	74.3	482	9	AI871469
C 11	342	73.2	386	10	AW594286
C 12	338	72.4	356	12	BQ024597
C 13	332	71.1	428	10	BF222471
C 14	324	69.4	437	9	AI933364
C 15	291	62.3	406	10	BE463932
C 16	285	61.0	612	9	AI886944
C 17	242	51.8	300	10	BF939932
C 18	242	51.8	334	9	AA304416
C 19	242	51.8	409	9	AI031859
C 20	241	51.6	396	9	AA083737
C 21	241	51.6	458	9	AA101878
C 22	238	51.0	403	9	AU156068
C 23	205	43.9	977	9	AL570794
C 24	201	43.0	271	9	AA367347
C 25	197	42.2	414	14	T27131
C 26	183	39.2	492	14	R60026
C 27	156	33.4	1200	9	AL571908
C 28	155	33.2	1201	9	AL570843
C 29	132	28.3	229	10	BF589966
C 30	127	27.2	528	13	EX118770
C 31	127	27.2	596	14	W67823
C 32	118	25.3	616	14	W91937
C 33	81	17.3	544	10	BF114974
C 34	80	17.1	885	14	CF456943
C 35	75	16.1	348	9	AA083736
C 36	61	13.1	557	9	AI871137
C 37	58	12.4	556	12	BM994192
C 38	52	11.1	640	12	BM992527
C 39	52	11.1	744	12	BQ006379
C 40	41	8.8	574	13	BQ234959
C 41	34	7.3	491	9	AI676745
C 42	30	6.4	972	29	CNS05DCI
C 43	29	6.2	311	9	AA363384
C 44	29	6.2	405	9	AA065331
C 45	29	6.2	423	29	LBAP096B01
C 46	29	6.2	482	14	T58423
C 47	29	6.2	732	12	BG328052
C 48	29	6.2	855	12	BG528108
C 49	28	6.0	323	12	BJ600203
C 50	28	6.0	345	10	AW509816
C 51	28	6.0	534	13	BY479938
C 52	28	6.0	572	10	BF266891
C 53	28	6.0	666	12	RJ580384
C 54	27	5.8	283	28	AZ057146
C 55	27	5.8	350	14	D21888
C 56	27	5.8	351	28	BH701978
C 57	27	5.8	393	14	CD343253
C 58	27	5.8	393	14	CD343259
C 59	27	5.8	449	9	AU062636
C 60	27	5.8	462	9	AA392984
C 61	27	5.8	462	10	AW746753
C 62	27	5.8	475	10	BE363676
C 63	27	5.8	486	10	AW678716
C 64	27	5.8	487	10	AW746812
C 65	27	5.8	497	9	AI520474
C 66	27	5.8	508	9	AA539308
C 67	27	5.8	509	10	BE122310
C 68	27	5.8	515	10	AW678672
C 69	27	5.8	525	14	CD814478
C 70	27	5.8	541	14	CD344181
C 71	27	5.8	554	12	BI241290
C 72	27	5.8	560	28	AZ778502
C 73	27	5.8	564	9	AA950806
C 74	27	5.8	572	9	AA950805
C 75	27	5.8	572	10	AW679994
C 76	27	5.8	575	14	CF416326
C 77	27	5.8	592	10	AW745375

```

c 78      27      5.8      627      10      AW678825
c 79      27      5.8      639      12      BI638657
c 80      27      5.8      708      14      CD825475
c 81      27      5.8      736      10      BE391757
c 82      27      5.8      818      28      CNS020XB
c 83      27      5.8      867      28      BH719045
c 84      27      5.8      893      29      CG963374
c 85      27      5.8      915      29      CNS030H2
c 86      27      5.8      916      29      CNS01Q7J
c 87      27      5.8      948      13      BU906591
c 88      27      5.8      1355      13      BU716281
c 89      26      5.6      207      10      BB023936
c 90      26      5.6      217      9      AJ284913
c 91      26      5.6      257      9      AI256810
c 92      26      5.6      279      14      CD49837
c 93      26      5.6      280      14      CD343152
c 94      26      5.6      284      10      BF479424
c 95      26      5.6      295      9      AV137522
c 96      26      5.6      301      9      AU255321
c 97      26      5.6      310      13      BY468581
c 98      26      5.6      316      14      CD343666
c 99      26      5.6      317      12      BJ475346
c 100     26      5.6      317      28      AZ006229

```

## ALIGNMENTS

```

RESULT 1
BM975075/c
LOCUS
DEFINITION
  BM975075 728 bp mRNA linear EST 20-FEB-2003
  UI-CF-EC1-acf-1-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
  UI-CF-EC1-acf-1-23-0-UI 3', mRNA sequence.
ACCESSION
  BM975075
VERSION
  BM975075.1 GI:19592666
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 728)
  Ronaldo, M.F., Lennon, G. and Soares, M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
  97044477
  8989548
  Contact: McCray, PB
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com) or from Open Biosystems
  (www.openbiosystems.com).
  The following repetitive elements were found in this cDNA
  sequence: 408-444, >(CAG)n#Simple repeat (885-728,
  >(GGA)n#Simple repeat (matched complement)
  Seq primer: M13_FORWARD
  POLYA=yes.

```

## FEATURES

```

source
  1..728
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="UI-CF-EC1-acf-1-23-0-UI"
  /tissue_type="Lung"

```

```

/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EC1"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaudo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGTGGCTTAC.
TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG LIB=UI-CF-EC1
TAG SEQ=AAGTGGCTTAC"
ORIGIN

```

Query Match 99.8%; Score 466; DB 12; Length 728;

Best Local Similarity 100.0%; Pred. No. 1.2e-221; Indels 0; Gaps 0;

Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 CAATGTTTGCCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGTGCTAAAGCTCTG 60
DB 484 CAATGTTTGCCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGTGCTAAAGCTCTG 425
QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGTCGGTGCCAGAAA 120
DB 424 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACTGTCGGTGCCAGAAA 365
QY 121 GTCTCTCTGCCACTGACGCCGCCCATCAGGATTTGGGCTTCTTTCCCTCTCTCTCTG 180
DB 364 GTCTCTCTGCCACTGACGCCGCCCATCAGGATTTGGGCTTCTTTCCCTCTCTCTG 305
QY 181 TGCTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 304 TGCTCTCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
QY 241 GAGAAAGTGGGGATGCTAAGAAAGCTGGAGATAGGGAACAGAGAGAGAGAGAGAG 300
DB 244 GAGAAAGTGGGGATGCTAAGAAAGCTGGAGATAGGGAACAGAGAGAGAGAGAG 185
QY 301 GGCTAGGGGGCTGCTCTATTTAAAGTGGTGTGTTATGATTTCTTATCTTATCTA 360
DB 184 GGCTAGGGGGCTGCTCTATTTAAAGTGGTGTGTTATGATTTCTTATCTTATCTA 125
QY 361 AGATATTAGGGCTGCTGCTCATTTAAGAAATTTGCCCTTCCCTGCTGCTGCTGCT 420
DB 124 AGATATTAGGGCTGCTGCTCATTTAAGAAATTTGCCCTTCCCTGCTGCTGCTGCT 65
QY 421 AGATGTTCTGCTGCTAATATGCTTTTATATAAACAGTAAAGCT 466
DB 64 AGATGTTCTGCTGCTAATATGCTTTTATATAAACAGTAAAGCT 19

```

## RESULT 2

AI149981/c

LOCUS

DEFINITION

AI149981

AI149981.1

GI:3678450

EST.

KEYWORDS

Source

ORGANISM

Homo sapiens

Human

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AI149981 450 bp mRNA linear EST 10-NOV-1998  
 qf38h10.x1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1752355  
 3' similar to contains element PTRS repetitive element ;, mRNA  
 sequence.

AI149981

AI149981.1

GI:3678450

EST.

KEYWORDS

Source

ORGANISM

Homo sapiens

Human

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 450)  
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb-remail.nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
 Insert length: 718 Std Error: 0.00  
 Seq primer: -40ml3 fwd. RT from Amersham  
 High quality sequence stop: 417.

Location/Qualifiers  
 1. 450  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1752355"  
 /sex="male"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares testis NHT"  
 /notes="Vector: p7f73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5']  
 TGTACCAATCTGAAGTGGGAGCGCGGCCAATTTTTTTTTTTT 3']  
 Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7f73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match	95.7%;	Score 447;	DB 9;	Length 450;
Best Local Similarity	100.0%;	Pred. No. 3.6e-212;		
Matches 447;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 21 CCCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTTAA 80

DB 450 CCCCCAAGCCCTTTACCTATGCTGCTGAACGCTGCTGCTGCTGCTGCTGCTTAA 391

QY 81 AGGCTCATGCTTGGAGTGGGAGCTGTCGGTGCACAGAAAGTCTCTTCTGCCACTGACGC 140

DB 390 AGGCTCATGCTTGGAGTGGGAGCTGTCGGTGCACAGAAAGTCTCTTCTGCCACTGACGC 331

QY 141 CCCCATCAGGATGGGCGCTTCTTTCCCGCTTCCTTTCTGTCCTCCTGCTCATCGGC 200

DB 330 CCCCATCAGGATGGGCGCTTCTTTCCCGCTTCCTTTCTGTCCTCCTGCTCATCGGC 271

QY 201 TGGCATGACCTGCAGGCCAAGCCCGTGGGGAAGGGAGAAAAGTGGGGATGGCTA 260

DB 270 TGGCATGACCTGCAGGCCAAGCCCGTGGGGAAGGGAGAAAAGTGGGGATGGCTA 211

QY 261 AGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTAGGGGGGCTGCCTAT 320

DB 210 AGAAAGCTGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTAGGGGGGCTGCCTAT 151

QY 321 TTAAGTGGTGTGTTTATGATCTTTACTAATTTATACAAAGTATTAGGGCCCTGTTCA 380

DB 150 TTAAGTGGTGTGTTTATGATCTTTACTAATTTATACAAAGTATTAGGGCCCTGTTCA 91

QY 381 TTAAGAAATGTTCCCTTCCCTGCTCAATGTTTGTAAAGATGCTTCTGTGTAAATAT 440

DB 90 TTAAGAAATGTTCCCTTCCCTGCTCAATGTTTGTAAAGATGCTTCTGTGTAAATAT 31

QY 441 GTCTTTTATAATACACAGTTTAAAGCTG 467

Subtraction: Two Approaches To Facilitate Gene Discovery.  
Genome Research 6, 791-806.  
TAG\_TISSUE=brain  
TAG\_LIB=NCI CGAP\_Brn23  
TAG\_SEO=ATATC"

## ORIGIN

```
Query Match      88.7%; Score 414; DB 9; Length 449;
Best Local Similarity 100.0%; Pred. No. 1.1e-195;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	54	54	431
Qy	GTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTGGAGTGGGGACTGTCGGTGC	113	
Db	GTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTGGAGTGGGGACTGTCGGTGC	372	

QY	114	CCAGAAAGTCTCTTCTGCGCACTGACGCGCCCCCATCAGGGATTGGCGATTGGCCCCCTTC	173
Dδ	371	CCAGAAAGTCTCTTCTGCGCACTGACGCGCCCCCATCAGGGATTGGCGATTGGCCCCCTTC	312

Qy	174	CTTCTGTGTCCTGCTCATCGGCGCTGCCATGACCTGCAGCCAAAGCCAGCCCCGTGG	233
Db	311	CTTCTGTGTCCTGCTCATCGGCGCTGCCATGACCTGCAGCCAAAGCCAGCCCCGTGG	252

Qy	234	GGAAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAAACAGAGAGGGTA	293
Db	251	GGAAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAAACAGAGAGGGTA	192

Qy	294	GTGGGTGGGCTAGGGGGCTGCCTATTATAAGTGGTGTGTTATGATCTTACTAATT	353
Db	191	GTGGGTGGGCTAGGGGGCTGCCTATTATAAGTGGTGTGTTATGATCTTACTAATT	132

QY 354 TATACAAGATATTAAAGCCCTGTTTCATTAGAAATTTTCCTTCCCTCGTGTTCATG 413

Dδ 131 TATACAAGATATTAAAGCCCTGTTTCATTAGAAATTTTCCTTCCCTCGTGTTCATG 72

Qy 414 TTTGTAAGAGTGTTCGTGTAAATATGCTTTATAATAACAGTTAAAAGCTG 467

Dδ 71 TTTGTAAGAGTGTTCGTGTAAATATGCTTTATAATAACAGTTAAAAGCTG 18

RESULT 4  
CA444588/C

LOCUS	CD44588	517 bp	linear	EST 08-NOV-2002
DEFINITION	UI-H-DT1-awl-m-06-0-UI-s1 NCI CGAP DT1 Homo sapiens CDNA clone			
ACCESSION	CD444588			

VERSION	CA444588.1	GI:24809008
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
NCI-CCGAP <http://www.nci.nih.gov/ccgap>  
1 (bases 1 to 517)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Robert Strausberg Ph.D.

Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Assayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
The following reagent(s) element(s) was/were used in this work:

```

sequence: 409-448, > (CAG)n#simple_repeat
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
FEATURES

```

```
source
1. .517
/organism="Homo sapiens"
/mol_type="mRNA"
```

```

/db_xref="taxon:9606"
/clone="UI-H-DT1-aw1-m-06-0-UI"
/tissue_type="Metastatic Chondrosarcoma"
/dev_stages="Adult"
/lab_host="DHI0B (Life Technologies)"
/clone_lib="NCI CGAP Drl"
/note="Organ: Lung; Vector: p77T3-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; NCI CGAP Drl is a normalized cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p77T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (drl)8 tail. The sequence tag for this library is AACTGTTCCG. TAG TISSUE=Lung metastatic chondrosarcoma TAG LIB=UI-H-DT1 TAG_SEQ=AACTGTTCCG"

```

## ORIGIN

```
Query Match      88.7%; Score 414; DB 14; Length 517;
Best Local Similarity 100.0%; Pred. No. 1.le-195;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	54	432
Qy	54	432
Db	54	432

Qy	114	372	Db
	CTGAAAGTCTCTTCTGCGCATGTGACGCCGCCCATCAGGGAATTGGCCCTTCTTCCCTTC	CAGAAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGGAATTGGCCCTTCTTCCCTTC	
	173	313	

Qy	174	312	253
D <sub>b</sub>	CTTCTGTGTCCTCTCATCGGCGTCCATGACCTGCAGCAAGCCAGCCCCGTGG	CTTCTGTGTCCTGCTCATCGGCTGCCATGACCTGCAGCAAGCCAGCCCCGTGG	CTTCTGTGTCCTGCTCATCGGCTGCCATGACCTGCAGCAAGCCAGCCCCGTGG

Qy	234	GGAGCGGAGAAAGTGGGGGATGGCTTAAGAAAGCTGGAGATAGGGAGCAGAAAGGGGTA	293
Db	252	GGAGCGGAGAAAGTGGGGGATGGCTTAAGAAAGCTGGAGATAGGGAGCAGAAAGGGGTA	193

Qy	294	GTGGGTGGGTAGGGGGCTGCTTATTTAAAGTGGTGTATTAATGATCTCTATACTAATT	353
Db	192	GTGGGTGGGTAGGGGGCTGCTTATTTAAAGTGGTGTATTAATGATCTCTATACTAATT	133

Qy 354 TATACAAGATATTATAGGCCCTGTTCAATTAGAAAATGTTCCCTTCCCTGCTTCAATG 413

Db 132 TATACAAGATATTATAGGCCCTGTTCAATTAGAAAATGTTCCCTTCCCTGCTTCAATG 73

Qy	414	TTTGTAAGATGTTCTCGTGAATAATGCTTTATATAAAACAGTTAAAAGCTG	467
Db	72	TTTGTAAGATGTTCTCGTGAATAATGCTTTATATAAAACAGTTAAAAGCTG	19

RESULT 5  
BQ006616/C

LOCUS	BQ006616	668 bp	linear	EST 26-MAR-2000
DEFINITION	U1-H-E11-aza-n-02-0-UI-s1 NCI CGAP_E11 Homo sapiens CDNA clone			
IMAGE	5846209 3',	mRNA sequence.		
ACCESSION	BQ006616			

VERSION BQ006616.1 GI:19731516  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 ORGANISM Homo sapiens

REFERENCES  
 AUTHORS  
 NIT-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
 1 (bases 1 to 668)  
 P-1mates; Catarrhini; Hominiidae; Homo.  
 Craniata; Vertebrata; Euteleostomi;  
 Chordata; Euthalia; Mammalia; Eutheria;  
 Metazoa; Eukaryota;

**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

```

JOURNAL      Tumor Gene Index
COMMENT      Unpublished (1997)
              Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: Dr. Jose Mercuende
              cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Dr. M. Bento Soares, University of Iowa
              through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
              The following repetitive elements were found in this cDNA
              sequence: 409-448, >(CAG)n#simple_repeat
              Seq primer: M13 FORWARD
              POLYA=Yes.

FEATURES     source
  source     Location/Qualifiers
    1..668   /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:5846209"
              /tissue_type="Chondrosarcoma"
              /dev_stage="Adult"
              /lab_host="DH10B (Life Technologies)"
              /clone_lib="NCI CGAP E11"
              /note="Organ: Left Pelvis; Vector: p7T3-Pac (Pharmacia)
              with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
              NCI CGAP E11 is a normalized cDNA library containing the
              following tissue(s): Chondrosarcoma. The library was
              constructed according to Bonaldo, Lennon and Soares,
              Genome Research, 6:791-806, 1996. First strand cDNA
              synthesis was primed with an oligo-dr primer containing a
              Not I site. Double stranded cDNA was ligated to an EcoR I
              adaptor, digested with Not I, and cloned directionally
              into p7T3-Pac vector. The oligonucleotide used to prime
              the synthesis of first-strand cDNA contains a library tag
              sequence that is located between the Not I site and the
              (dT)18 tail. The sequence tag for this library is
              ACATTGCAC.
              TAG TISSUE=chondrosarcoma
              TAG LIB=UT-H-E11
              TAG_SEQ=ACATTGCAC"

ORIGIN
Query Match      98.7%; Score 414; DB 12; Length 668;
Best Local Similarity 100.0%; Pred. No. 1.2e-195;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTGGAGTGGGACTGCTGGTGC 113
DB 432 GCTGCTGCTGCTGCTGCTGCTTAAGGCTCATGCTGGAGTGGGACTGCTGGTGC 373
QY 114 CCAGAAAGTCTTCTTCCCACTGACGCCCCCATCAGGGATTGGSCCTTCTTCCCTTC 173
DB 372 CCAGAAAGTCTTCTTCCCACTGACGCCCCCATCAGGGATTGGSCCTTCTTCCCTTC 313
QY 174 CTTTCTGTCTCTGCTCATGCGCTGCGCATGACCTGAGCCAGCCAGCCAGCCGCTGG 233
DB 312 CTTTCTGTCTCTGCTCATGCGCTGCGCATGACCTGAGCCAGCCAGCCAGCCGCTGG 253
QY 234 GGAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGAGACAGAGAGGTA 293
DB 252 GGAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGAGACAGAGAGGTA 193
QY 294 GTGGGTGGGCTAGGGGGCTGCTCTATTATTAAGTGGTGTGTTATGATCTTATCTAATT 353
DB 192 GTGGGTGGGCTAGGGGGCTGCTCTATTATTAAGTGGTGTGTTATGATCTTATCTAATT 133
QY 354 TATCAAGAATATTAAGGCCCTGTTCTATTAAAGAAATGTTCCCTTCCCTGTGTCAATG 413
DB 132 TATCAAGAATATTAAGGCCCTGTTCTATTAAAGAAATGTTCCCTTCCCTGTGTCAATG 73
QY 414 TTTCTAAGATTGTTCTGTGAATATGTTCTTATATAAAGCTGTTAAAGCTG 467

```

---

```

Db 72 TTTCTAAGATTGTTCTGTGAATATGTTCTTATATAAAGCTGTTAAAGCTG 19

RESULT 6
BE855577/c
LOCUS
DEFINITION
  BE855577 555 bp mRNA linear EST 29-SEP-2000
  7910e01.x1 NCI CGAP Brn23 Homo sapiens cDNA clone IMAGE:3306072 3'
  similar to contains element OFR repetitive element ;, mRNA
  sequence.
  BE855577
ACCESSION
  BE855577.1 GI:10367753
VERSION
  BE855577.1
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE
    1 (bases 1 to 555)
    NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute / National Institute of Neurological
    Disorders and Stroke, Brain Tumor Genome Anatomy Project
    (CGAP/BTGP), Tumor Gene Index
  JOURNAL
    Unpublished (1998)
  COMMENT
    Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-remail.nih.gov
    Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
    Ph.D.
    cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
    Bonaldo, Ph.D.
    cDNA Library Arrayed by: Greg Lemmon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone Distribution: NCI-CGAP clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL, send email to:
    info@image.llnl.gov
    Seq primer: -40UP from Gibco
    High quality sequence stop: 472.

FEATURES     source
  source     Location/Qualifiers
    1..555   /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:3306072"
              /tissue_type="glioblastoma (pooled)"
              /lab_host="DH10B"
              /clone_lib="NCI CGAP Brn23"
              /note="Organ: Brain; Vector: p7T3D-Pac (Pharmacia) with a
              modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
              strand cDNA was primed with a Not I oligo(dT) primer [5'
              TGTTCAATCTGAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTTT
              T 3']; double-stranded cDNA was ligated to Eco RI
              adaptors (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of the modified p7T3 vector.
              Library is normalized, and was constructed by Bento
              Soares and M. Fatima Bonaldo."

ORIGIN
Query Match      82.2%; Score 384; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 1.1e-180;
Matches 384; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTGCTATCCACCTCCCAAGCCCTTTACCTATGCTGTGCTAACGCTGCTG 60
DB 468 CAATGTTGCTATCCACCTCCCAAGCCCTTTACCTATGCTGTGCTAACGCTGCTG 409
QY 61 CTGCTGCTGCTGCTTAAGGCTCATGCTGGAGTGGGACTGCTGGTGCCTCCAGAA 120
DB 408 CTGCTGCTGCTGCTTAAGGCTCATGCTGGAGTGGGACTGCTGGTGCCTCCAGAA 349
QY 121 GTCTCTTCTGCCACTGACGCCCCCATCAGGAGTTGGCCCTTCTTCCCTTCTCTTCTG 180
DB 348 GTCTCTTCTGCCACTGACGCCCCCATCAGGAGTTGGCCCTTCTTCCCTTCTCTTCTG 289
QY 181 TGTCTCTGCTCATCGGCTGCGCATGACCTGACGCAAGCCAGCCCGCTGGGAGGG 240

```

```

Db      288  TGTCTCCTGCTCATGGCGCTGCCATGACCTGACGCCAAGCCAGCCCGCGTGGGAAAGG 229
QY      241  GAGAAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAAAGAGGGTAGTGGGTG 300
Db      228  GAGAAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAAAGAGGGTAGTGGGTG 169
QY      301  GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATCTTATCTAATTTATACAA 360
Db      168  GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATCTTATCTAATTTATACAA 109
QY      361  AGATATTAAAGCCCTGTTTCATTAA 384
Db      108  AGATATTAAAGCCCTGTTTCATTAA 85

RESULT 7
BX434223/c
LOCUS   BX434223 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE009YJ24
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION BX434223
VERSION   BX434223.1 GI:30777248
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 889)
AUTHORS  Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE    Full-length cDNA libraries and normalization
JOURNAL  Unpublished (2001)
COMMENT  Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqreg@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10757.r For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAK025CH09NM1&cluster=10757.r. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0BAK025CH09NM1.

FEATURES             source
    Location/Qualifiers
        ..889
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0DE009YJ24"
        /tissue_type="PLACENTA"
        /clone_lib="Homo sapiens PLACENTA"
        /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
        with a NotI-oligo(dT) primer. Five prime end enriched,
        double-strand cDNA was digested with Not I and cloned into
        the Not I and EcoRV sites of the pCMVSPORT 6 vector.
        Library was not normalized."

ORIGIN
Query Match          79.9%; Score 373; DB 13; Length 889;
Best Local Similarity 99.8%; Pred. No. 3.8e-175;
Matches 423; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  CAATGTTTGCTATCCACCTCCCAAGCCCTTACATGCTGCTGCTGAACGCTGCTG 60
Db      462  CAATGTTTGCTATCCACCTCCCAAGCCCTTACATGCTGCTGCTGAACGCTGCTG 403
QY      61  CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGACTGGTCGGTCCAGAAA 120
Db      402  CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGACTGGTCGGTCCAGAAA 343
QY      121  GTCTCTTCTGCACTGACGCGCCCACTCAGGAGATTGGGCTTCTTTCCCTCTCTTCTG 180
Db      342  GTCTCTTCTGCACTGACGCGCCCACTCAGGAGATTGGGCTTCTTTCCCTCTCTTCTG 283

```

```

QY      181  TGTCTCCTGCTCATGGCGCTGCCATGACCTGACGCCAAGCCAGCCCGCGTGGGAAAGG 240
Db      282  TGTCTCCTGCTCATGGCGCTGCCATGACCTGACGCCAAGCCAGCCCGCGTGGGAAAGG 223
QY      241  GAGAAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAAAGAGGGTAGTGGGTG 300
Db      222  GAGAAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAAAGAGGGTAGTGGGTG 163
QY      301  GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATCTTATCTAATTTATACAA 360
Db      162  GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATCTTATCTAATTTATACAA 103
QY      361  AGATATTAAAGCCCTGTTTCATTAAAGAAATTTGCCCTTCCCTCTGTTCAATGTTGTAA 420
Db      102  AGATATTAAAGCCCTGTTTCATTAAAGAAATTTGCCCTTCCCTCTGTTCAATGTTGTAA 43
QY      421  AGAT 424
Db      42  AGAT 39

RESULT 8
AA099445 431 bp mRNA linear EST 11-MAY-1997
LOCUS   ZK85f07.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
DEFINITION IMAGE:489637 5', similar to contains element PTR5 repetitive element
ACCESSION AA099445
VERSION   AA099445.1 GI:1645393
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS  Hillier,J., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,S., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Mortis,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.,
and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
JOURNAL   97044478
MEDLINE   8889549
PUBMED    8889549
COMMENT   Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 528 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham.

FEATURES             source
    Location/Qualifiers
        1..431
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="GDB:3804224"
        /db_xref="taxon:9606"
        /clone="IMAGE:489637"
        /sex="female"
        /dev_stage="adult"
        /lab_host="DH10B"
        /clone_lib="Soares_pregnant_uterus_NbHPU"
        /note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;
        Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
        oligo(dT) primer [5'
        AACTGGAGAAATTCGCGCGCTTTTTTTTTTTTTTTT 3'],
        double-stranded cDNA was ligated to Eco RI adaptors

```

(Pharmacacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

```

ORIGIN
Query Match          79.2%; Score 370; DB 9; Length 431;
Best Local Similarity 99.8%; Pred. No. 1.1e-173;
Matches 420; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 CTATCCACTCCCAAGCCCTTTACCTATCTGCTGCTACGCTGCTGCTGCTGCTG 70
DB 10 CTATCCACTCCCAAGCCCTTTACCTATCTGCTGCTGCTGCTGCTGCTGCTG 69
QY 71 TCTCTCTTAAAGGCTCATGCTTGGAGTGGGACTGCTGGTCCGACAGAAAGTCTTCTG 130
DB 70 TCTCTCTTAAAGGCTCATGCTTGGAGTGGGACTGCTGGTCCGACAGAAAGTCTTCTG 129
QY 131 CCACTGAGCCGCCCATCAGGAGTGGGCTTCTTCCCTCTTCTTCTGCTGCTGCTG 190
DB 130 CCACTGAGCCGCCCATCAGGAGTGGGCTTCTTCCCTCTTCTTCTGCTGCTGCTG 189
QY 191 CTCATCGGCTGCCATGACCTGCGACCAAGCCAGCCCGCTGGGAGAGGAGAAAGTGG 250
DB 190 CTCATCGGCTGCCATGACCTGCGACCAAGCCAGCCCGCTGGGAGAGGAGAAAGTGG 249
QY 251 GGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGGGTGGCTAGGGGG 310
DB 250 GGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGGGTGGCTAGGGGG 309
QY 311 GCTGCTTATTTAAAGTGGTGTATGATCTTATCTAATTTATACAAAGATTTAAG 370
DB 310 GCTGCTTATTTAAAGTGGTGTATGATCTTATCTAATTTATACAAAGATTTAAG 369
QY 371 GCCCTGTTTCAATTAAGAAATGTTCCCTCCCTGTTCAATGTTTGAAGATTTGTTCT 430
DB 370 GCCCTGTTTCAATTAAGAAATGTTCCCTCCCTGTTCAATGTTTGAAGATTTGTTCT 429
QY 431 G 431
DB 430 G 430

```

```

RESULT 9
CA438256/c
LOCUS
DEFINITION
  UI-H-DT1-avw-g-05-0-UI-s1 NCI CGAP DT1 Homo sapiens cDNA clone
  UI-H-DT1-avw-g-05-0-UI 3', mRNA sequence.
ACCESSION
  CA438256
VERSION
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 887)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
  from Dr. M. Bento Soares, bento-soares@uiowa.edu
  The following repetitive elements were found in this cDNA
  sequence: 409-448, >(CAG)n$Simple_repeat
Seq primer: M13 FORWARD
POLYA-Yes
Location/Qualifiers

```

## source

```

1. 687
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-DT1-avw-g-05-0-UI"
/tissue_type="Metastatic Chondrosarcoma"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP DT1"
/note="Organ: Lung; Vector: p773-Pac (Pharmacacia) with a
modified polylinker; Site_1: EcoR I; Site 2: Not I;
NCI CGAP DT1 is a normalized cDNA library containing the
following tissue(s): Metastatic Chondrosarcoma in Lung.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into p773-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is AACTGTTGG.
TAG_TISSUE=lung metastatic chondrosarcoma
TAG_LIB=UI-H-DT1
TAG_SEQ=AACTGTTGG"

```

## ORIGIN

```

Query Match          77.7%; Score 363; DB 14; Length 687;
Best Local Similarity 99.8%; Pred. No. 3.6e-170;
Matches 413; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 54 GCTCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACGTGGTGGTGC 113
DB 432 GCTCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTGGAGTGGGACGTGGTGGTGC 373
QY 114 CCAGAAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGAGTGGGCTTCTTCCCTTC 173
DB 372 CCAGAAAGTCTCTTCTGCCACTGACGCCGCCCATCAGGAGTGGGCTTCTTCCCTTC 313
QY 174 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 233
DB 312 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253
QY 234 GGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGAAACAGAGGGTA 293
DB 252 GGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGAAACAGAGGGTA 193
QY 294 GTGGGTGGCTAGGGGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGG 353
DB 192 GTGGGTGGCTAGGGGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGG 133
QY 354 TATACAAAGATTTAAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTCAATG 413
DB 132 TATACAAAGATTTAAGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTCAATG 73
QY 414 TTTGTAAGATTTGTTCTGCTGTAATATGCTTTATATAAACAGTAAAGCTG 467
DB 72 TTTGTAAGATTTGTTCTGCTGTAATATGCTTTTATATAAACAGTAAAGCTG 19

```

## RESULT 10

```

AI871469/c
LOCUS
DEFINITION
  AI871469 x1 NCI CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2429969 3',
  similar to contains element TARI repetitive element ;, mRNA
  sequence.
ACCESSION
  AI871469
VERSION
  AI871469.1 GI:5545518
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1409 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 450.

FEATURES
source
1..482
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2429969"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn25"
/organ="Brain; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAATGGAGCGCCGACATAGTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 74.3%; Score 347; DB 9; Length 482;
Best Local Similarity 99.8%; Pred. No. 3.4e-162;
Matches 467; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 CAATGTTTGCTATCCACCTCCCGACGCCCTTACCTATGCTGCTAACGCTGCTG 60
468 CAATGTTTGCTATCCACCTCCCGACGCCCTTACCTATGCTGCTAACGCTGCTG 409
61 CTGCTGCTGCTGCTATCAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCAGAA 120
408 CTGCTGCTGCTGCTGCTATCAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCAGAA 349
121 GTCTCTTCTGCTGCTGCTATCAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCCTTCTG 180
348 GTCTCTTCTGCTGCTGCTATCAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCCTTCTG 289
181 TGCTCTGCTGCTGCTGCTATCAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCAGAGG 240
288 TGCTCTGCTGCTGCTGCTATCAAGGCTCATGCTGGAGTGGGACTGCTGCTGCCAGAGG 229
241 GAG-AAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTGGGT 299
228 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGGGTAGTGGGT 169
300 GGGCTAGGGGGCTGCTTATTAAGTGGTGGTATGATTTCTTATTAATTAATA 359
168 GGGCTAGGGGGCTGCTTATTAAGTGGTGGTATGATTTCTTATTAATTAATA 109
360 AAGATATTAAGGCCCTGCTTATTAAGAAATGTTCCCTTCCCTGCTTCAATGTTTGT 419
108 AAGATATTAAGGCCCTGCTTATTAAGAAATGTTCCCTTCCCTGCTTCAATGTTTGT 49

ORIGIN
Query Match 73.2%; Score 342; DB 10; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.1e-159;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

87 ATGCTTGGAGTGGGACTGCTCGGTGCCAGAAAGTCTCTTCTGCCACTGACGCCCCCAT 146
386 ATGCTTGGAGTGGGACTGCTCGGTGCCAGAAAGTCTCTTCTGCCACTGACGCCCCCAT 327
147 CAGGGATTTGGGCTTCTTTCCCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 206
326 CAGGGATTTGGGCTTCTTTCCCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 267
207 GACCTCAGCCAGCCCGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 266
266 GACCTCAGCCAGCCCGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 207
267 CTGGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGGGCTGCTTATTTAAAG 356
206 CTGGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGGGCTGCTTATTTAAAG 147

```

```

QY 327 TGGTGTGTTATGATCTTATCTAATTAATTAACAAGATATTAAAGCCCTGTTCAATTAAGA 386
Db 146 TGGTGTGTTATGATCTTATCTAATTAATTAACAAGATATTAAAGCCCTGTTCAATTAAGA 87

QY 387 AATTGTTCCCTTCCCTGTTGTTCAATGTTTGAAGATTGTT 428
Db 86 AATTGTTCCCTTCCCTGTTGTTCAATGTTTGAAGATTGTT 45

RESULT 12
BQ024597/c
LOCUS
DEFINITION
  UI-1-BB1p-aus-e-04-0-UI.s1 NCI CGAP P16 Homo sapiens cDNA clone
  UI-1-BB1p-aus-e-04-0-UI 3', mRNA sequence.
ACCESSION
  BQ024597
VERSION
  BQ024597.1
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  1. GI:19759876
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 356)
REFERENCE
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
JOURNAL
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. Steven Brown
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be obtained
  from Dr. M. Bento Soares, bento-soares@uiowa.edu
  Seq primer: M13 FORWARD
  POLYA=Yes.
FEATURES
  source
  1. .356
  /location="Qualifiers"
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="UI-1-BB1p-aus-e-04-0-UI"
  /tissue_type="Placenta"
  /dev_stage="Full Term"
  /lab_host="DH10B (Life Technologies)"
  /clone_lib="NCI CGAP P16"
  /note="Organ: Placenta; Vector: pTT73-Pac (Pharmacia) with
  a modified polylinker; Site 1: EcoR I; Site 2: Not I;
  NCI CGAP P16 is a subtracted cDNA library constructed
  according to Bonaldo, Lennon and Soares, Genome Research,
  6:791-806, 1996. First strand cDNA synthesis was primed
  with an oligo-dT primer containing a Not I site. Double
  stranded cDNA was ligated to an EcoR I adaptor, digested
  with Not I, and cloned directionally into pTT73-Pac
  vector. The oligonucleotide used to prime the synthesis of
  first-strand cDNA contains a library tag sequence that is
  located between the Not I site and the (dT)18 tail. The
  sequence tags for this library are GA, AGGAA. For
  additional information, contact: Bento Soares,
  bento-soares@uiowa.edu
  TAG TISSUE=placenta human full term
  TAG LIB=UI-1-BB1p
  TAG_SEQ=AGGAA"
ORIGIN

Query Match 72.4%; Score 338; DB 12; Length 356;
Best Local Similarity 100.0%; Pred. No. 1e-157;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 GCCACTGAGCCCCCATCAGGATGGGCCCTCTTCCCTCTCTTCTGTCCTGTCCTG 189
Db 356 GCCACTGAGCCCCCATCAGGATGGGCCCTCTTCTCCCTCTCTTCTGTCCTGTCCTG 297

```

```

QY 190 CCTCATCGCCCTGCCATGACCTGCAGCCCAAGCCCGTGGGAGCGGAGAAAGTG 249
Db 296 CCTCATCGCCCTGCCATGACCTGCAGCCCAAGCCCGTGGGAGCGGAGAAAGTG 237

QY 250 GGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGG 309
Db 236 GGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTAGTGGGTGGGCTAGGGG 177

QY 310 GGCTGCCCTATTAAAGTGGTGTGTTATGATCTTATCTAATTAACAAGATATTAA 369
Db 176 GGCTGCCCTATTAAAGTGGTGTGTTATGATCTTATCTAATTAACAAGATATTAA 117

QY 370 GGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTCTGTTCAATGTTGTAAAGATTGTT 429
Db 116 GGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTCTGTTCAATGTTGTAAAGATTGTT 57

QY 430 TGTGTAAATATGTTCTTTATATAATAACACAGTAAAGCTG 467
Db 56 TGTGTAAATATGTTCTTTATATAATAACAGTAAAGCTG 19

RESULT 13
BQ222471/c
LOCUS
DEFINITION
  hr75g11.x1 NCI CGAP Kid11 Homo sapiens cDNA clone IMAGE:3134372 3',
  similar to contains element OFR repetitive element ;, mRNA
  sequence.
ACCESSION
  BQ222471
VERSION
  BQ222471.1
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  1. GI:11129648
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
JOURNAL
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  cDNA Library Preparation: M. Bento Soares, Ph.D.
  cDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone Distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL, send email to:
  info@image.llnl.gov
  High quality sequence stop: 416.
FEATURES
  source
  1. .428
  /location="Qualifiers"
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:3134372"
  /lab_host="DH10B"
  /clone_lib="NCI CGAP Kid11"
  /note="Organ: Kidney; Vector: pTT73D-Pac (Pharmacia) with
  a modified polylinker; Site 1: Not I; Site 2: Eco RI;
  Plasmid DNA from the normalized library NCI CGAP Kid3
  prepared, and ss circles were made in vitro. Following HAP
  purification, this DNA was used as tracer in a subtractive
  hybridization reaction. The driver was PCR-amplified cDNAs
  from a pool of 5,000 clones made from the same library
  (cloneIDs 1322376-1323911, 1456007-1456775, and
  1500552-1502855). Subtraction by Bento Soares and M.
  Fatima Bonaldo."
ORIGIN

Query Match 71.1%; Score 332; DB 10; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.1e-154;

```





```

/clone lib="NCI CGAP Kid11"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
hybridization, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 132376-132391, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

```

## ORIGIN

```

Query Match 62.3%; Score 291; DB 10; Length 406;
Best Local Similarity 99.7%; Pred. No. 3.3e-134;
Matches 341; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 126 TTCTGCCACTGACGCCCCCATCAGGATGGGCTTCTTCCCTTCTTCTGTGCT 185
Db |||||
QY 186 COTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGAGAA 245
Db |||||
QY 285 CTTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGAGAA 226
QY 246 AGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTA 305
Db |||||
QY 225 AGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTGGGCTA 166
QY 306 GGGGGCTGCGCTTATTAAAGTGGTGTATGATTTCTTACTAATTTATACAAAGATA 365
Db |||||
QY 165 GGGGGCTGCGCTTATTAAAGTGGTGTATGATTTCTTACTAATTTATACAAAGATA 106
QY 366 TTAAGGCGCTGTTCAATTAAGAAATGTTCCCTTCCCTGTGTTCAATTTGTAAGATT 425
Db |||||
QY 105 TTAAGGCGCTGTTCAATTAAGAAATGTTCCCTTCCCTGTGTTCAATTTGTAAGATT 46
QY 426 GTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 467
Db |||||
QY 45 GTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 4

```

```

RESULT 16
AI886944/c
LOCUS
DEFINITION
w194d03.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2432549 3',
mRNA sequence.
AI886944
AI886944.1 GI:5592108
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 517 Std Error: 0.00
Seq primer: -40UP from Gibco

```

```

FEATURES
source
High quality sequence stop: 312.
Location/Qualifiers
1..612

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2432549"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Brn25"

```

```

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCATAGGTTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

```

## ORIGIN

```

Query Match 61.0%; Score 285; DB 9; Length 612;
Best Local Similarity 100.0%; Pred. No. 3.4e-131;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 TCTCTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGGA 242
Db |||||
QY 287 TCTCTGCTCATCGGCTGCGCATGACCTGCAGCCAGCCAGCCCGTGGGAGGGGGA 228
Db |||||
QY 243 GAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTGGG 302
Db |||||
QY 227 GAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGTAGTGGGTGGG 168
QY 303 CTAGGGGGCTGCGCTTATTAAAGTGGTGTATGATTTCTTACTAATTTATACAAAG 362
Db |||||
QY 167 CTAGGGGGCTGCGCTTATTAAAGTGGTGTATGATTTCTTACTAATTTATACAAAG 108
QY 363 ATATTAGGCGCTGTTCAATTAAGAAATGTTCCCTTCCCTGTGTTCAATTTGTAAG 422
Db |||||
QY 107 ATATTAGGCGCTGTTCAATTAAGAAATGTTCCCTTCCCTGTGTTCAATTTGTAAG 48
QY 423 ATTGTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 467
Db |||||
QY 47 ATTGTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 3

```

```

RESULT 17
BF939932/c
LOCUS
DEFINITION
nac64d01.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3439129
3', mRNA sequence.
BF939932
BF939932.1 GI:12357252
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
Cloning Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 517 Std Error: 0.00
Seq primer: -40UP from Gibco

```



```

REFERENCE 1 (bases 1 to 409)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-remail@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 921 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 379.
FEATURES
Location/Qualifiers
1..409
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1668815"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn23"
/notes="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 51.8%; Score 242; DB 9; Length 409;
Best Local Similarity 100.0%; Pred. No. 1e-109;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 226 CCCGCGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGA 285
Db 247 CCCGCGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGA 188
Qy 286 AGAGGGTAGTGGGTGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTTA 345
Db 187 AGAGGGTAGTGGGTGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTTA 128
Qy 346 TACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGT 405
Db 127 TACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGT 68
Qy 406 GTTCAATGTTTGAAGAATGTTCTGTGTAATATGCTTTTATATAAAGTAAAGC 465
Db 67 GTTCAATGTTTGAAGAATGTTCTGTGTAATATGCTTTTATATAAAGTAAAGC 8
Qy 466 TG 467
Db 7 TG 6
RESULT 20
AA083737 396 bp mRNA linear EST 21-OCT-1996
LOCUS zn39d09.s1 StrataGene endothelial cell 937223 Homo sapiens cDNA
DEFINITION clone IMAGE:549809 3', mRNA sequence.
ACCESSION AA083737
VERSION AA083737.1 GI:1625813
KEYWORDS EST.

```

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 396)
AUTHORS Hillier L., Lennon G., Becker M., Bonaldo M.F., Chiapelli B.,
Chisoe S., Dietrich N., DuBuque T., Pavello A., Gish W.,
Hawkins M., Hultman M., Kucaba T., Lacy M., Le M., Le N.,
Mardis E., Moore B., Morris M., Parsons J., Frange C., Rifkin L.,
Rohlfing T., Scheinberg K., Soares M.B., Tan F., Thierry-Mieg J.,
Trevaskis E., Underwood K., Wohlmann P., Waterston R., Wilson R.
and Marra M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd from Amersham.
FEATURES
Location/Qualifiers
1..396
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3928326"
/db_xref="taxon:9606"
/clone="IMAGE:549809"
/dev_stage="umbilical vein, 1 passage"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene endothelial cell 937223"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Umbilical vein endothelial cells, passaged once. Average
insert size: 1.0 Kb; Uni-ZAP XR Vector; ~5' adaptor
sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5'
CTCGAGTCTTTTCTTTTCTTTTCTTTT 3'"
ORIGIN
Query Match 51.6%; Score 241; DB 9; Length 396;
Best Local Similarity 100.0%; Pred. No. 3.1e-109;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 227 CCCGCGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAA 286
Db 245 CCCGCGGGGAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAA 186
Qy 287 GAGGGTAGTGGGTGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 346
Db 185 GAGGGTAGTGGGTGGGCTAGGGGGCTGCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 126
Qy 347 ACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGTG 406
Db 125 ACTAATTTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGTG 66
Qy 407 TTCAATGTTTGAAGAATGTTCTGTGTAATATGCTTTTATATAAAGTAAAGCT 466
Db 65 TTCAATGTTTGAAGAATGTTCTGTGTAATATGCTTTTATATAAAGTAAAGCT 6
Qy 467 G 467
Db 5 G 5
RESULT 21
AA101878 458 bp mRNA linear EST 11-MAY-1997
LOCUS zk85f07.s1 Soares pregnant uterus_NBHPU Homo sapiens cDNA clone
DEFINITION IMAGE:489637 3', mRNA sequence.

```

```

ACCESSION      AA101878
VERSION        AA101878.1  GI:1645281
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens

REFERENCE
AUTHORS        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
                Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
                Chissoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W.,
                Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
                Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
                Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
                Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
                and Marra, M.
TITLE          Generation and analysis of 280,000 human expressed sequence tags
JOURNAL        Genome Res. 6 (9), 807-828 (1996)
MEDLINE        97044478
PUBMED        889549
COMMENT        Contact: Wilson RK
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: est@watson.wustl.edu
                This clone is available royalty-free through LLNL; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Insert Length: 528  Std Error: 0.00
                Seq primer: -40M13 fwd. from Amersham.
                Location/Qualifiers
FEATURES       source
               1..458
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="GDB:3804224"
                /db_xref="taxon:9606"
                /clone="IMAGE:489637"
                /sex="female"
                /dev_stage="adult"
                /lab_host="DH10B"
                /clone_lib="Soares_pregnant_uterus_NbHPV"
                /note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;
                Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
                oligo(dT) primer [5'
                AACTGGAAGATTCGGCGCCCTTTTCTTTTCTTTT 3'],
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. Library
                went through one round of normalization. Library
                constructed by M. Fatima Bonaldo."

ORIGIN
Query Match      51.6%; Score 241; DB 9; Length 458;
Best Local Similarity 100.0%; Pred. No. 3.2e-109;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CCGTGGGAGGGGAGAAAGTGGGGATGCTTAAGAAAGCTGGGAGATAGGGAACAGAA 286
      |||
Db 252 CCGTGGGAGGGGAGAAAGTGGGGATGCTTAAGAAAGCTGGGAGATAGGGAACAGAA 193
      |||
QY 287 GAGGTAGTGGTGGCTAGGGGGCTGCCTTATTAAAGTGGTGTATTGATCTTAT 346
      |||
Db 192 GAGGTAGTGGTGGCTAGGGGGCTGCCTTATTAAAGTGGTGTATTGATCTTAT 133
      |||
QY 347 ACTAATTTATCAAGATATTAAGGCCCTGTTCATTAAAGAAATGTTCCTTCCCTGTG 406
      |||
Db 132 ACTAATTTATCAAGATATTAAGGCCCTGTTCATTAAAGAAATGTTCCTTCCCTGTG 73
      |||
QY 407 TTCAATGTTGTAAGATGTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCT 466
      |||
Db 72 TTCAATGTTGTAAGATGTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCT 13
      |||
QY 467 G 467

RESULT 23
AL570794/c
LOCUS      AL570794      977 bp      mRNA      linear      EST 31-MAY-2003
DEFINITION      AUI56068 PLACE1 Homo sapiens CDNA clone PLACE1002140 3', mRNA
                sequence.
ACCESSION      AUI56068
VERSION        AUI56068.1  GI:11017589
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
                Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
                1 (bases 1 to 403)
                Oca, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
                Yamamoto, J., Wakamatsu, A., Ozawa, M., Nakamura, Y., Nagai, T.,
                Sugano, S. and Isogai, T.
                HRI human CDNA project (Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S.,
                Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Ozawa, M.,
                Nakamura, Y., Nagai, T., Sugano, S., Isogai, T.)
                Unpublished (2000)
                Contact: Takao Isogai
                Genomics Laboratory
                Helix Research Institute
                1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
                Tel: 81-438-52-3975
                Fax: 81-438-52-3986
                Email: genomics@hri.co.jp
                HRI human CDNA project; 5' - & 3'-end one pass sequencing: Helix
                Research Institute; cDNA library construction: Department of
                Virology, Institute of Medical Science, University of Tokyo, and
                Helix Research Institute.
                Location/Qualifiers
FEATURES       source
               1..403
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="PLACE1002140"
                /tissue="placenta"
                /clone_lib="PLACE1"
                /note="Vector: pME18SFL3"

ORIGIN
Query Match      51.0%; Score 238; DB 9; Length 403;
Best Local Similarity 99.7%; Pred. No. 1e-107;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 TGTGTCTCTGCTCATCGGCTGCCATGACCTGCAGCCCAAGCCCGCCGCGGGAAG 238
      |||
Db 289 TGTGTCTCTGCTCATCGGCTGCCATGACCTGCAGCCCAAGCCCGCCGCGGGAAG 230
      |||
QY 239 GCGAGAAAGTGGGGATGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGG 298
      |||
Db 229 GCGAGAAAGTGGGGATGCTTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGG 170
      |||
QY 299 TGGGCTAGGGGGGCTGCCTTATTAAAGTGGTGTATTGATCTTATATACTAATTATAC 358
      |||
Db 169 TGGGCTAGGGGGGCTGCCTTATTAAAGTGGTGTATTGATCTTATATACTAATTATAC 110
      |||
QY 359 AAAGATATTAAAGCCCTGTTCATTAAAGAAATGTTCCTTCCCTGCTGTTCATGTTGT 418
      |||
Db 109 AAAGATATTAAAGCCCTGTTCATTAAAGAAATGTTCCTTCCCTGCTGTTCATGTTGT 50
      |||
QY 419 AAAGATTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 467
      |||
Db 49 AAAGATTTCTGTGTAATATGCTTTTATATAAACAAGTTAAAGCTG 1
      |||

```



<p> <b>TITLE</b> The WashU-Merck EST Project  <b>UNPUBLISHED</b> (1995)  <b>CONTACT:</b> Wilson RK  Washington University School of Medicine  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  <b>TEL:</b> 314 286 1800  <b>FAX:</b> 314 286 1810  <b>EMAIL:</b> est@watson.wustl.edu  <b>HIGH QUALITY SEQUENCE STOPS:</b> 355  <b>SOURCE:</b> IMAGE Consortium, LLNL  This clone is available royalty-free through LLNL ; contact the  IMAGE Consortium (info@image.llnl.gov) for further information.  <b>SEQ PRIMER:</b> SP6  <b>HIGH QUALITY SEQUENCE STOP:</b> 363.  Location/Qualifiers  1. .492  /organism="Homo sapiens"  /mol_type="mRNA"  /db_xref="GDB:415555"  /db_xref="taxon:9606"  /clone="IMAGE:43014"  /sex="female"  /dev_stage="73 days post natal"  /lab_host="PH10B (ampicillin resistant)"  /clone_lib="Soares infant brain LNIP"  /notes="Organ: whole brain; Vector: Lfamid BA; Site 1: Not  I; Site 2: Hind III; 1st strand cDNA was primed with a Not  I - oligo(dT) primer 15',  AACTGAAGAATCGCGCGCGGAGGAATTTTATTTT 3';  double-stranded cDNA was ligated to Hind III adaptors  (Pharmacia), digested with Not I and directionally cloned  into the Not I and Hind III sites of the Lfamid BA vector.  Library went through one round of normalization. Library  constructed by Bento Soares and M. Patina Bonaldo." </p>	
<p> <b>ORIGIN</b>  Query Match 39.2%; Score 183; DB 14; Length 492;  Best Local Similarity 99.6%; Pred. No. 3.1e-80;  Matches 233; Conservative 0; Mismatches 1; Indels 0; Gaps 0;    QY 227 CCGCTGGGAGGAGGAAGTGGGGGATGGCTTAAGAAAGCTGGGAGATAGGACAGAA 286  DB 252 CCGCTGGGAGGAGGAGGAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGACAGAA 193  QY 287 GAGGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGCTTTATGATTCCTAT 346  DB 192 GAGGGTAGTGGGTGGCTAGGGGGCTGCCTTATTTAAAGTGGCTTTATGATTCCTAT 133  QY 347 ACTAATTTATACAAAGATATTAAGGCCCTGTTCATTAAAGAAATGTTCCCTTCCCTGTG 406  DB 132 ACTAATTTATACAAAGATATTAAGGCCCTGTTCATTAAAGAAATGTTCCCTTCCCTGTG 73  QY 407 TTCAATGTTTGAAGATGTTCTCTGTAAATATGCTTTATAATAACAGTTA 460  DB 72 TTCAATGTTTGAAGATGTTCTCTGTAAATATGCTTTATAATAACAGTTA 19 </p>	
<p> <b>RESULT 27</b>  AL571908/c  <b>LOCUS</b> AL571908 Homo sapiens FLACENTA COT 25-NORMALIZED Homo sapiens CDNA  <b>DEFINITION</b> clone CSODI029YL01 3-PRIME, mRNA sequence.  <b>ACCESSION</b> AL571908  <b>VERSION</b> AL571908.2 GI:31293299  <b>KEYWORDS</b> EST.  <b>SOURCE</b> Homo sapiens (human)  <b>ORGANISM</b> Homo sapiens  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  1. (bases 1 to 1200)  Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  Full-length cDNA libraries and normalization  Unpublished (2001) </p>	



```

/lab host="DH10B"
/clone_lib="NCI_CGAP_Kid11"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI CGAP Kid3 was
prepared; and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

```

## ORIGIN

```

Query Match      28.3%; Score 132; DB 10; Length 229;
Best Local Similarity 99.5%; Pred. No. 8.7e-55;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      285 AAGAGGCTAGTGGCTAGGGGGCTGCTTATTAAAGTGGTGTATTATGATCTT 344
      |||||
Db       188 AAGAGGCTAGTGGCTAGGGGGCTGCTTATTAAAGTGGTGTATTATGATCTT 129
      |||||
QY      345 ATACTAATTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTCCCTG 404
      |||||
Db       128 ATACTAATTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTCCCTG 69
      |||||
QY      405 TGTTCATGTTTGAAGATGTTCTGTGTAATATGCTTTATATAAACAGTTAAAG 464
      |||||
Db        68 TGTTCATGTTTGAAGATGTTCTGTGTAATATGCTTTATATAAACAGTTAAAG 9
      |||||
QY      465 CTG 467
      |||||
Db        8 CTG 6

```

## RESULT 30

```

BX118770
LOCUS      BX118770 Soares_fetal_heart_NBH19W Homo sapiens cDNA clone
DEFINITION IMAG998G06783 ; IMAGE:342941, mRNA sequence.
ACCESSION BX118770
VERSION    BX118770.1 GI:27882396
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
Radlof,U., Schneider,D. and Korn,B.
Human UnigeneSet - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAG998G06783.
RZPDLIB: I.M.A.G.E. cDNA Clone Collection;
Human UnigeneSet - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de

```

This clone is available royalty-free from RZPD;

contact RZPD (clone@rzpd.de) for further information. Seq primer:

M13r, Primer sequence: TTTCACACAGGAACAGCTATGAC.

## FEATURES

## source

```

1..528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

```

/clone="IMAG998G06783 ; IMAGE:342941"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares_fetal_heart_NBH19W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGAGCGGCATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."

```

## ORIGIN

```

Query Match      27.2%; Score 127; DB 13; Length 528;
Best Local Similarity 100.0%; Pred. No. 3e-52;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      341 TCTTACTAATTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCC 400
      |||||
Db       1 TCTTACTAATTATACAAAGATATTAAAGCCCTGTTTCATTAAAGAAATGTTCCCTTCC 60
      |||||
QY      401 CCTGTGTTCAAGTGTGTAAGATGTTCTGTGTAATATGCTTTTATAAACAGTTA 460
      |||||
Db        61 CCTGTGTTCAAGTGTGTAAGATGTTCTGTGTAATATGCTTTTATAAACAGTTA 120
      |||||
QY      461 AAGCTG 467
      |||||
Db       121 AAGCTG 127

```

## RESULT 31

```

W67823
LOCUS      W67823 zd38f03.r1 Soares_fetal_heart_NBH19W Homo sapiens cDNA clone
DEFINITION IMAG998G06783 ; IMAGE:342941 5', mRNA sequence.
ACCESSION W67823
VERSION    W67823.1 GI:1376742
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Huitman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 630 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 394.
Location/Qualifiers
1..596
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1268316"
/db_xref="taxon:9606"

```

## FEATURES

## source

```

1..596
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1268316"
/db_xref="taxon:9606"

```



```

/clone="IMAGE:342941"
/sex="unknown"
/dev stage="19 weeks"
/lab host="DH10B (ampicillin resistant)"
/clone lib="Soares fetal heart NBH19W"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCACTGAAGTGGAGCGCCGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."

ORIGIN
Query Match      27.2%; Score 127; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 3.1e-52;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 TCTTATCTACTAATTACAAAGATATTAGGCCCTGTTCAATTAAGAAATGTTCCCTCC 400
    |||||||
Db 1 TCTTATCTAATTATACAAAGATATTAGGCCCTGTTCAATTAAGAAATGTTCCCTCC 60

QY 401 CTTGTTTCAATGTTTGAAGATGTTCTGTGTAATATGTCCTTTATAATAACAGTTA 460
    |||||||
Db 61 CTTGTTTCAATGTTTGAAGATGTTCTGTGTAATATGTCCTTTATAATAACAGTTA 120

QY 461 AAAGCTG 467
    |||||||
Db 121 AAAGCTG 127

RESULT 32
W91937
LOCUS      zh47c08.r1 Soares fetal liver spleen INFLS_S1 Homo sapiens cDNA
DEFINITION clone IMAGE:415214 5', mRNA sequence.
ACCESSION W91937
VERSION    W91937.1 GI:1424298
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 616)
AUTHORS    Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
            Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
            and Marra,M.
            Generation and analysis of 280,000 human expressed sequence tags
            Genome Res. 6 (9), 807-828 (1996)
            97044478
            PUBMED 8889549
COMMENT    Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 720 Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 451.
            Location/Qualifiers
            1. .616

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDE:1323684"
/db_xref="taxon:9606"
/clone="IMAGE:415214"
/sex="male"
/dev stage="20 week-post conception fetus"
/lab host="DH10B (ampicillin resistant)"
/clone lib="Soares fetal liver spleen INFLS_S1"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
This is a subcloned version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AATCGGAAGAAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN
Query Match      25.3%; Score 118; DB 14; Length 616;
Best Local Similarity 100.0%; Pred. No. 9.8e-48;
Matches 118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 AATTATACAAGATATTAAAGCCCTGTTCAATTAAGAAATGTTCCCTCCCTGTGTC 409
    |||||||
Db 1 AATTATACAAGATATTAAAGCCCTGTTCAATTAAGAAATGTTCCCTCCCTGTGTC 60

QY 410 AATGTTTGAAGATGTTCTGTGTAATATGTCCTTTATAATAACAGTTAAAGCTG 467
    |||||||
Db 61 AATGTTTGAAGATGTTCTGTGTAATATGTCCTTTATAATAACAGTTAAAGCTG 118

RESULT 33
BF114974
LOCUS      hr72a04.x1 NCI-CGAP Kid11 Homo sapiens cDNA clone IMAGE:3133998 3',
DEFINITION mRNA sequence.
ACCESSION BF114974
VERSION    BF114974.1 GI:10984450
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 544)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
JOURNAL
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 468.
            Location/Qualifiers
            1. .544
FEATURES
            source
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:3133998"
            /lab host="DH10B"
            /clone lib="NCI CGAP Kid11"
            /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

```

a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NC1\_CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo.

## ORIGIN

Query Match 17.3%; Score 81; DB 10; Length 544;  
Best Local Similarity 100.0%; Pred. No. 3e-29;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 387 AATGTTCCCTCCCTGCTGTCATGTTGTAAGATTGTCGTGTAATATGCTTT 446  
Db 544 AATGTTCCCTCCCTGCTGTCATGTTGTAAGATTGTCGTGTAATATGCTTT 485  
QY 447 ATAATAACAGTTAAAGCTG 467  
Db 484 ATAATAACAGTTAAAGCTG 464

RESULT 34  
CF456943  
LOCUS 885 bp mRNA linear EST 04-SEP-2003  
DEFINITION AGENCOURT 15337517 Human Anterior Horn Homo sapiens cDNA clone  
IMAGE:30514432 5', mRNA sequence.

ACCESSION CF456943  
VERSION CF456943.1 GI:34456599  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 885)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

## AUTHORS

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)

Tissue Procurement: Dr. James R. Lupski

cDNA Library Preparation: Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDAM577 row: 0 column: 17

High quality sequence start: 4

High quality sequence stop: 598.

## FEATURES

## source

1..885  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30514432"  
/tissue\_type="Peripheral Nervous system"  
/lab\_host="DH10B (T1 phage-resistant)"  
/clone\_lib="Human Anterior Horn"  
/note="Vector: pCMV-SPORT5.1; Site\_1: EcoRV (destroyed);  
Site\_2: NotI; Library is oligo-dT primed and directionally  
cloned (EcoRV site is destroyed upon cloning). Average  
insert size 2.1 kb. Library was constructed by  
Invitrogen."

## ORIGIN

Query Match 17.1%; Score 80; DB 14; Length 885;  
Best Local Similarity 100.0%; Pred. No. 1e-28;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAATGTTGCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGAACGCTGTG 60  
Db 624 CAATGTTGCTATCCACCTCCCAAGCCCTTTACCTATGCTGCTGAACGCTGTG 683  
QY 61 CTGCTGCTGCTGCTGCTTAA 80  
Db 684 CTGCTGCTGCTGCTGCTTAA 703

RESULT 35  
AA083736  
LOCUS 348 bp mRNA linear EST 21-OCT-1996  
DEFINITION zn39409.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA  
clone IMAGE:549809 5', mRNA sequence.

ACCESSION AA083736

VERSION AA083736.1 GI:1625812

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 348)

## AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

889549

## COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 324.

## FEATURES

## source

1..348  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GD3:3928326"  
/db\_xref="taxon:9606"  
/clone="IMAGE:549809"  
/dev\_stage="umbilical vein, 1 passage"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene endothelial cell 937223"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT.  
Umbilical vein endothelial cells, passaged once. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor  
sequence: 5' GAATTCGGCAG 3' ~3' adaptor sequence: 5'  
CTCGAGTTTATTTTTTTT 3'"

## ORIGIN

Query Match 16.1%; Score 75; DB 9; Length 348;  
Best Local Similarity 100.0%; Pred. No. 2.8e-26;  
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 160 TTCTTTCCCTTCCTTTCTGTCCTCGCTCAFCGCGCTGCCATGACCTGACGCCAA 219  
Db 147 TTCTTTCCCTTCCTTTCTGTCCTCGCTCAFCGCGCTGCCATGACCTGACGCCAA 206  
QY 220 GCCCAGCCCCGTGGG 234

ACCESSION	BM994192	BM994192.1	GI:19719093	
VERSION	EST.			
KEYWORDS				
SOURCE	Organism	Human sapiens		
ORGANISM		Human sapiens (human)		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS		1 (bases 1 to 556)		
TITLE		NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT		Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: ccgaps-remail.nih.gov Tissue Procurement: Dr. Jose Mercuende cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Seq primer: M13 FORWARD POLYA=Yes.		
FEATURES		Location/Qualifiers		
source		1..556		
		/organism="Homo sapiens"		
		/mol_type="mRNA"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:5865900"		
		/tissue_type="Metastatic Chondrosarcoma"		
		/dev_stage="Adult"		
		/lab_host="DH10B (Life Technologies)"		
		/clone_lib="NCI_CGAP_DTO"		
		/note="Organ: Lung; Vector: pWT3-Pac (Pharmacia) with a modified polylinker; Site: Ecor I; Site 2: Not I; NCI_CGAP DTO is a cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pWT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTCGG.		
ORIGIN		TAG TISSUE=Lung metastatic chondrosarcoma TAG_LIB=UT-H-DTO TAG_SEQ=AACTGTCGG		
Query Match		12.4%; Score 58; DB 12; Length 556;		
Best Local Similarity		100.0%; Pred. No. 9.3e-18;		
Matches		58; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
QY	410	AATGTTTGTAAAGATTCTTCGTGTAATATGCTTTATATAACAGTTAAAGCTG	467	
Db	537	AATGTTTGTAAAGATTCTTCGTGTAATATGCTTTATATAACAGTTAAAGCTG	480	
RESULT 38				
BM992527/c				
LOCUS		640 bp mRNA linear EST 17-JUN-2000		
DEFINITION		UI-H-DTO-att-k-01-0-UI.sl NCI_CGAP DTO Homo sapiens cDNA clone		
		IMAGE:5864194 3', mRNA sequence.		
ACCESSION		BM992527		
VERSION		BM992527		
KEYWORDS		EST.		
SOURCE		BM992527.1	GI:19711916	
ORGANISM		Human sapiens (human)		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE		1 (bases 1 to 640)		

**AUTHORS**  
**TITLE** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**JOURNAL** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
**COMMENT** Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
**Tissue Procurement:** Dr. Jose Mercuende  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 The following repetitive elements were found in this cDNA  
 sequence: 1-21, >AT rich#Low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

**FEATURES**  
 source  
 1. .640  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5864184"  
 /tissue\_type="Metastatic Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP DT0"  
 /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site: 1: Ecor I; Site 2: Not I; NCI CGAP DT0 is a cDNA library containing the following tissue(s): Metastatic Chondrosarcoma in Lung. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AACTGTCGG.  
 TAG TISSUE=lung metastatic chondrosarcoma  
 TAG LIB=UI-H-DT0  
 TAG\_SEQ=AACTGTCGG"

**ORIGIN**  
 Query Match 11.1%; Score 52; DB 12; Length 640;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-15;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TGTAAAGATTGTTCTGTTAAATATGCTTTATATAACAGTTAAAGCTG 467  
 |||||  
 Db 617 TGTAAAGATTGTTCTGTTAAATATGCTTTATATAACAGTTAAAGCTG 566  
 |||||

**RESULT 39**  
 BQ006379/c  
 LOCUS  
 DEFINITION UI-H-E11-aza-i-11-UI.s1 NCI CGAP\_E11 Homo sapiens CDNA clone  
 IMAGE:5846038 3', mRNA sequence.  
 ACCESSION BQ006379  
 VERSION BQ006379.1 GI:19731279  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.  
 1 (bases 1 to 744)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

**Tissue Procurement:** Dr. Jose Mercuende  
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 The following repetitive elements were found in this cDNA  
 sequence: 1-21, >AT rich#Low\_complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

**FEATURES**  
 source  
 1. .744  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5846098"  
 /tissue\_type="Chondrosarcoma"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP E11"  
 /notes="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site: 1: Ecor I; Site 2: Not I; NCI CGAP E11 is a normalized cDNA library containing the following tissue(s): Chondrosarcoma. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is ACACTGCAC.  
 TAG TISSUE=chondrosarcoma  
 TAG LIB=UI-H-E11  
 TAG\_SEQ=ACACTGCAC"

**ORIGIN**  
 Query Match 11.1%; Score 52; DB 12; Length 744;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-15;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 373 CCTGTTCAATAGAAATGTTCCCTTCCCTGTTCAATGTTGTAAGAT 424  
 |||||  
 Db 660 CCTGTTCAATAGAAATGTTCCCTTCCCTGTTCAATGTTGTAAGAT 609  
 |||||

**RESULT 40**  
 BQ234959  
 LOCUS  
 DEFINITION BQ234959.g1 Canis Total Brain cDNAs Canis familiaris CDNA clone  
 hd52b09 5', mRNA sequence.  
 ACCESSION BQ234959  
 VERSION BQ234959.1 GI:20430835  
 KEYWORDS EST.  
 SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 1 (bases 1 to 574)  
 O'Shaughnessy, A.L., Palmer, L., McCombie, W.R., Baker, J.P., Bahret, A., Cunniss, D., Dedhia, N., de la Bastide, M., Katzenberger, F., King, L., Kirchoff, K.A., Miller, B., Muller, S., Nascimento, L.U., Preston, R.R., Shah, R.S., Spiegel, L.A., Zutavern, T., Santos, L. and Hannon, G.J.  
 Expressed sequence tags from Canis familiaris (dog) (5\_2002)  
 Unpublished (2002)  
 Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884

Fax: 516 367 8874  
Email: mccombie@cshl.org  
Plate: hd52 row: b column: 09  
Seq primer: -21M13UnivRev  
High quality sequence stop: 574.  
Location/Qualifiers  
1. .574  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/db\_xref="taxon:9615"  
/clone="hd52b09"  
/clone\_lib="Canis Total Brain cDNAs"  
/note="Vector: Lambda Zap II; The library was provided by Greg Hannon and Lee Santos (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stratagene zap cDNA synthesis kit. It was made from dog whole brain cells. please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."

ORIGIN

Query Match 8.8%; Score 41; DB 13; Length 574;  
Best Local Similarity 100.0%; Pred. No. 2.9e-09;  
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TATGATCTTATCTACTAATTTATACAAAGATATTAAGGCCCT 375  
|||||  
Db 395 TATGATCTTATCTACTAATTTATACAAAGATATTAAGGCCCT 435

RESULT 41  
LOCUS AI676745/c 491 bp mRNA linear EST 19-MAY-1999  
DEFINITION etnEST0524 Eth1 Eimeria tenella cDNA clone etmJ085 5', mRNA  
sequence.  
ACCESSION AI676745.1 GI:4877225  
VERSION AI676745  
KEYWORDS EST.  
SOURCE Eimeria tenella  
ORGANISM Eimeria tenella  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.

REFERENCE  
1 (bases 1 to 491)  
Wan,K.L., Chong,S.P., Ng,S.T., Tomley,F.M. and Jangi,M.S.  
Survey of genes in Eimeria tenella merozoite  
Unpublished (1999)  
Contact: Wan KL  
Centre for Gene Analysis and Technology  
Universiti Kebangsaan Malaysia  
43600 UKM Bangi, Selangor DE, Malaysia  
Tel: 6 03 8292997  
Fax: 6 03 8293249  
Email: klawan@pkriscc.cc.ukm.my  
PCR Primers  
FORWARD: T3  
BACKWARD: T7  
Seq primer: SK.

FEATURES  
source  
location/Qualifiers  
1. .491  
/organism="Eimeria tenella"  
/mol\_type="mRNA"  
/strain="Houghton"  
/db\_xref="taxon:S802"  
/clone="etmJ085"  
/dev\_stage="second generation merozoite"  
/lab\_host="XLI-Blue MRF/"  
/clone\_lib="Eth1"  
/note="Vector: Lambda ZAPII; Site\_1: EcoRI; Site\_2: XhoI;  
Second generation merozoites of E. tenella H were purified  
by column chromatography and mRNA extracted using a  
FASTTRACK kit (Invitrogen). cDNA was synthesised and a  
Uni-ZAP XR library was constructed using cDNA synthesis  
kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III  
Gold cloning kit (Stratagene). The library was amplified

once through E. coli XLI-Blue MRF'."

ORIGIN

Query Match 7.3%; Score 34; DB 9; Length 491;  
Best Local Similarity 100.0%; Pred. No. 9e-06;  
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 TGCTGCTAACGCTGCTGCTGCTGCTGCTGCTGCT 77  
|||||  
Db 486 TGCTGCTAACGCTGCTGCTGCTGCTGCTGCTGCT 453

RESULT 42  
LOCUS CNS05DCI 972 bp DNA linear GSS 01-SEP-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone  
042N02 of library A from Tetraodon nigroviridis, genomic survey  
sequence.  
ACCESSION AL332235  
VERSION AL332235.1 GI:8225993  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE  
1  
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence  
Nat. Genet. 25 (2), 235-238 (2000)

2  
20296633  
PUBMED 10835645

3 (bases 1 to 972)  
Genoscope.  
Direct Submission  
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers  
1. .972  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="042N02"  
/clone\_lib="A"  
/note="Genoscope sequence ID : COAA042DG01A1-end : T3"

ORIGIN

Query Match 6.4%; Score 30; DB 29; Length 972;  
Best Local Similarity 100.0%; Pred. No. 0.00098;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GCTAACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77  
|||||  
Db 687 GCTAACGCTGCTGCTGCTGCTGCTGCTGCT 716

```

RESULT 43
AA363384
LOCUS
DEFINITION
  311 bp mRNA linear EST 21-APR-1997
  sapiens Hypothetical protein CTG-B33 (GB:LI0376), mRNA sequence.
ACCESSION
AA363384
VERSION
AA363384.1 GI:2015704
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkie,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Melligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
TITLE
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL
Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE
96026280
PUBMED
7566098
COMMENT
Other ESTs: THCL72576
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
  1..311
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="ATCC (inhost):167640"
  /db_xref="taxon:9606"
  /sex="female"
  /dev_stage="adult, 71 yrs"
  /clone_libs="Ovary 1"
  /note="Organ: ovary; Vector: pBluescript SK-; Site_1:
  EcoRI; Site_2: XhoI"
ORIGIN
Query Match 6.2%; Score 29; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 195 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 223
FEATURES
source
  1..311
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="ATCC (inhost):167640"
  /db_xref="taxon:9606"
  /sex="female"
  /dev_stage="adult, 71 yrs"
  /clone_libs="Ovary 1"
  /note="Organ: ovary; Vector: pBluescript SK-; Site_1:
  EcoRI; Site_2: XhoI"
ORIGIN
Query Match 6.2%; Score 29; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 195 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 223

```

```

DEFINITION
f06500r Testis 5 Homo sapiens cDNA clone f06500 3' end, mRNA
sequence.
ACCESSION
AA065331
VERSION
AA065331.1 GI:1929211
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 405)
Guellaen,G.
Guellaen,G. Unpublished (1996)
Unpublished (1996)
Contact: Guellaen G
Unite INSERM 99
INSERM
Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
Tel: (33)149813530
Fax: (33)14980908
Email: guellaen@infobiogen.fr
This sequence derives from a clone which was selected from the cDNA
library - Testis 5 - using a repeat of 14 CAG as probe
Seq primer: M13 reverse
FEATURES
Location/Qualifiers
  1..405
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="f06500"
  /clone_lib="Testis 5"
  /note="Vector: pSPORT1; Site 1: MluI; Site 2: NotI; mRNA
  was prepared from human testis of a 27 years old man. cDNA
  was prepared using a 15mer oligo dt anchored by two
  degenerated bases at its 3'end and containing a NotI site
  at its 5'end. The cDNA was cloned between SalI and NotI
  sites of pSPORT1. The MluI-SalI fragment come from the
  adaptor used for the cloning. The 3' end is at the NotI
  site. cDNA corresponding to abundant species were
  eliminated from this library."
ORIGIN
Query Match 6.2%; Score 29; DB 9; Length 405;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
Db 281 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 309
RESULT 45
LBFA096B01
LOCUS
DEFINITION
  Leishmania braziliensis GSS, clone LBFA096B01, genomic survey
  sequence.
ACCESSION
BX541495
VERSION
BX541495.1 GI:32139660
KEYWORDS
GSS; genomic survey sequence.
SOURCE
Leishmania braziliensis
ORGANISM
Leishmania braziliensis
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania; Leishmania braziliensis species complex.
REFERENCE
1
Laurentino,E.C., Ruiz,J.C. and Cruz,A.K.
GSS analysis of the Leishmania braziliensis genome
Unpublished
2 (bases 1 to 423)
Cruz,A.K.
Direct Submission
Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo,
Department of Molecular and Cell Biology, FMRP, Avenida
Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL
Clone requests: akcruz@fmrp.usp.br.
COMMENT

```

```

FEATURES
  source
    1. .423
      /organism="Leishmania braziliensis"
      /mol_type="genomic DNA"
      /strain="MHOM/BR/75/M2904"
      /db_xref="taxon:5660"
      /clone="LBAF096B01"

ORIGIN
  Query Match
  Best Local Similarity 100.0%; Score 29; DB 29; Length 423;
  Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 49 CTAACGCTGCTGCTGCTGCTGCTGCTGCT 77
  Db 180 CTAACGCTGCTGCTGCTGCTGCTGCTGCT 208

RESULT 46
LOCUS T58423 482 bp mRNA linear EST 09-FEB-1995
DEFINITION YB60H09.r1 Stragatene ovary (#937217) Homo sapiens cDNA clone
IMAGE:75617 5', mRNA sequence.
ACCESSION T58423
VERSION T58423.1 GI:660260
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,M.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevasakis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PubMed 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 613
High quality sequence stops: 224
Clone is available royalty-free through LNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert Length: 613 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 224.

FEATURES
  source
    1. .482
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="GDB:505346"
      /db_xref="taxon:9606"
      /clone="IMAGE:75617"
      /sex="female"
      /dev_stage="49 year old"
      /lab_host="SOLR cells (kanamycin resistant)"
      /clone_lib="Stratagene ovary (#937217)"
      /note="Organ: ovary; Vector: Bluescript SK; Site 1: EcoRI;
      Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
      Total ovary tissue, normal, caucasian. Average insert
      size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
      GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
      CTCGAGTTTATTTTATTTT 3'"

ORIGIN
  Query Match
  Best Local Similarity 100.0%; Score 29; DB 12; Length 732;
  Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
  Db 473 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 501

RESULT 48
LOCUS BG528108 855 bp mRNA linear EST 03-APR-2001
DEFINITION 60255748F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:466572 5',
mRNA sequence.
ACCESSION BG528108
VERSION BG528108.1 GI:13519645
KEYWORDS EST.

FEATURES
  source
    1. .732
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:4547111"
      /tissue_type="adenocarcinoma cell line"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_15"
      /note="Organ: colon; Vector: pOT7; Site 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dT priming. Directionally
      cloned into EcoRI/XhoI sites using the following 5',
      adaptor: GGCACGAG(G). Size-selected >500bp for average
      insert size 1.8kb. Library constructed by Ling Hong in
      the laboratory of Gerald M. Rubin (University of
      California, Berkeley) using ZAP-cDNA synthesis kit
      (Stratagene) and Superscript II RT (Life Technologies)"

ORIGIN
  Query Match
  Best Local Similarity 100.0%; Score 29; DB 12; Length 732;
  Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
  Db 473 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 501

RESULT 47
LOCUS BG328052 732 bp mRNA linear EST 27-FEB-2001
DEFINITION 602427694F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4547111 5',
mRNA sequence.
ACCESSION BG328052
VERSION BG328052.1 GI:13134490
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 732)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: NIH Intramural Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1234 row: 1 column: 24
High quality sequence stop: 630.

FEATURES
  source
    1. .732
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:4547111"
      /tissue_type="adenocarcinoma cell line"
      /lab_host="DH10B (phage-resistant)"
      /clone_lib="NIH_MGC_15"
      /note="Organ: colon; Vector: pOT7; Site 1: XhoI; Site 2:
      EcoRI; cDNA made by oligo-dT priming. Directionally
      cloned into EcoRI/XhoI sites using the following 5',
      adaptor: GGCACGAG(G). Size-selected >500bp for average
      insert size 1.8kb. Library constructed by Ling Hong in
      the laboratory of Gerald M. Rubin (University of
      California, Berkeley) using ZAP-cDNA synthesis kit
      (Stratagene) and Superscript II RT (Life Technologies)"

ORIGIN
  Query Match
  Best Local Similarity 100.0%; Score 29; DB 12; Length 732;
  Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  QY 54 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 82
  Db 473 GCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 501

RESULT 48
LOCUS BG528108 855 bp mRNA linear EST 03-APR-2001
DEFINITION 60255748F1 NIH_MGC_59 Homo sapiens cDNA clone IMAGE:466572 5',
mRNA sequence.
ACCESSION BG528108
VERSION BG528108.1 GI:13519645
KEYWORDS EST.

```

```

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 855)
JOURNAL     NIH-MGC http://mgc.nci.nih.gov/.
COMMENT     National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L10M1497 row: 0 column: 21
            High quality sequence stop: 594.
FEATURES    Location/Qualifiers
            1..855
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4686572"
                /tissue type="mucoepidermoid carcinoma"
                /lab host="DH10B (T1 phage-resistant)"
                /clone lib="NIH_MGC_59"
                /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
                SfiI (ggccgccggcc); Site 2: SfiI (ggccattatggcc);
                Double-stranded cDNA was prepared from cell line RNA. 5'
                and 3' adaptors were used in cloning as follows: 5'
                adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
                sequence: 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)BN-3'
                (where B = A, C, or G and N = A, C, G, or T). Average
                insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                contained inserts by PCR. This library was enriched for
                full-length clones and was constructed by Clontech
                Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                Library."
ORIGIN
Query Match      6.2%; Score 29; DB 12; Length 855;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAAG 82
Db 483 GCTGCTGCTGCTGCTGCTGCTGCTTAAG 511

RESULT 49
LOCUS      BJ600203/c
DEFINITION BJ600203 323 bp mRNA linear EST 22-OCT-2003
            Bu600203 normalized full length cDNA library, chloronemata,
            caulonemata and rhizoid-like protonemata Physcomitrella patens
            subsp. patens cDNA clone pphn28014 3', mRNA sequence.
ACCESSION  BJ600203
VERSION     BJ600203.1 GI:37842195
KEYWORDS    EST.
SOURCE      Physcomitrella patens subsp. patens
ORGANISM    Physcomitrella patens subsp. patens
            Eukaryota; Viridiplantae; Streptophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1 (bases 1 to 323)
AUTHORS     Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H.,
            Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K.,
            Kohara, Y. and Hasebe, M.
            Comparative genomics of Physcomitrella patens gametophytic
            transcription and Arabidopsis thaliana: implication for land plant
            evolution
            Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
            22709184

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE       1 (bases 1 to 855)
JOURNAL     NIH-MGC http://mgc.nci.nih.gov/.
COMMENT     National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L10M1497 row: 0 column: 21
            High quality sequence stop: 594.
FEATURES    Location/Qualifiers
            1..855
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:4686572"
                /tissue type="mucoepidermoid carcinoma"
                /lab host="DH10B (T1 phage-resistant)"
                /clone lib="NIH_MGC_59"
                /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
                SfiI (ggccgccggcc); Site 2: SfiI (ggccattatggcc);
                Double-stranded cDNA was prepared from cell line RNA. 5'
                and 3' adaptors were used in cloning as follows: 5'
                adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
                sequence: 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)BN-3'
                (where B = A, C, or G and N = A, C, G, or T). Average
                insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
                contained inserts by PCR. This library was enriched for
                full-length clones and was constructed by Clontech
                Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                Library."
ORIGIN
Query Match      6.0%; Score 28; DB 12; Length 323;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 ACCTGCTGCTGCTGCTGCTGCTGCTGCTTA 79
Db 278 ACCTGCTGCTGCTGCTGCTGCTGCTGCTTA 251

RESULT 50
LOCUS      AW509816
DEFINITION Gae6a05.y1 Moss EST library PPU Physcomitrella patens cDNA clone
            PEP SOURCE ID: PPU110509 5', similar to TR:Q9ZVB7 Q9ZVB7 PUTATIVE
            STRICTOSIDINE SYNTHASE.; mRNA sequence.
ACCESSION  AW509816
VERSION     AW509816.1 GI:7147894
KEYWORDS    EST.
SOURCE      Physcomitrella patens
ORGANISM    Physcomitrella patens
            Eukaryota; Viridiplantae; Streptophyta; Bryophyta;
            Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
REFERENCE   1 (bases 1 to 345)
AUTHORS     Quatrano, R., Bashardes, S., Cove, D., Cuming, A., Knight, C.,
            Clifton, S., Marra, M., Hallier, L., Pape, D., Martin, J., Wylie, T.,
            Underwood, K., Theising, B., Allen, M., Bowers, Y., Person, B.,
            Swaller, T., Steptoe, M., Gibbons, M., Harvey, N., Ritter, E.,
            Jackson, Y., McCann, R., Waterston, R. and Wilson, R.
            Leeds/Wash U Moss EST Project
            Unpublished (1999)
            Contact: Ralph Quatrano
            Leeds/Wash U Moss EST Project
            Washington University School of Medicine

```



Wed May 26 09:35:15 2004

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Libraries were constructed by Dr. Stavros Bashirades as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco

High quality sequence stop: 293.

# FEATURES

source

```

1. .345
Location/Qualifiers
/organism="Physcomitrella patens"
/mol_type="mRNA"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID:PPU110509"
/tissue_type="protonemata: 7 day old tissue ammonium-grown"
/lab_host="DH10B"
/clone_lib="Moss EST library PPU"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Construction of the cDNA library was carried out using Stratagene's UniZAP- cDNA synthesis kit'. cDNA was constructed using an oligo dt primer/linker that contains a XhoI site within it. Following ds cDNA synthesis, EcoRI adapters were ligated to the blunt ends and sample was digested with XhoI. The result is cDNA with an EcoRI sticky end on one side and a XhoI sticky end on the other. This cDNA was ligated directionally in UniZAP arms. The vector is designed containing the pBluescript sequence as well as lambda DNA and cDNA is cloned within this pBluescript sequence. The vector was then packaged using Gold giga packaging extracts. Library was grown in XL1Blue MRF' cells and amplified. The library was excised by mass excision using Stratagene's Mass excision kit' that uses exassit as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids. SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered by using Qiaagen Midi prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation."

```

# ORIGIN

Query Match 6.0%; Score 28; DB 10; Length 345;  
 Best Local Similarity 100.0%; Pred.No. 0.0086;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 ACCTGCTGCTGCTGCTGCTGCTTA 79  
 |||||  
 Db 295 ACCTGCTGCTGCTGCTGCTGCTTA 322

Search completed: May 25, 2004, 19:19:03  
 Job time : 2889 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 15:14:46 ; Search time 95 Seconds  
(without alignments)

2728.022 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctatccacct.....taataaacagctaaagctg 467

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	467	100.0	50000	4	US-09-146-053-4
C 2	40.4	8.7	7218	1	US-08-232-463-14
C 3	38.2	8.2	4262	4	US-09-521-511C-10
C 4	37	7.9	2521	3	US-09-115-446-1
C 5	37	7.9	2521	1	US-09-565-590-1
C 6	37	7.9	2526	3	US-09-115-446-5
C 7	37	7.9	2526	4	US-09-565-590-5
C 8	36.4	7.8	3001	4	US-09-539-333D-178
C 9	35.8	7.7	800	1	US-08-468-853-7
C 10	35.8	7.7	800	1	US-08-468-855-7
C 11	35.8	7.7	800	1	US-08-310-357-7
C 12	35.8	7.7	800	1	US-08-468-852-7
C 13	35.8	7.7	800	2	US-08-468-857-7
C 14	35.6	7.6	107	4	US-09-205-995-75
C 15	35.6	7.6	169	4	US-09-205-995-72
C 16	35.6	7.6	190	4	US-09-205-995-77
C 17	35.4	7.6	430	4	US-09-621-976-16656
C 18	35.4	7.6	1853	1	US-08-553-110-2
C 19	35.4	7.6	4168	4	US-09-266-225D-17
C 20	35.4	7.6	4279	3	US-09-041-886-22
C 21	35.2	7.5	2160	3	US-09-588-256-1
C 22	35.2	7.5	3311	4	US-09-367-891A-5
C 23	35.2	7.5	9053	4	US-09-976-594-306
C 24	35.2	7.5	10807	1	US-08-206-176-7
C 25	35.2	7.5	10807	2	US-08-756-506-5
C 26	35.2	7.5	176373	3	US-09-128-155-17
C 27	35	7.5	10348	2	US-08-457-273B-41

35	7.5	10348	3	US-08-556-419-13	Sequence 13, Appl
35	7.5	10348	3	US-09-041-886-14	Sequence 1, Appl
35	7.5	10366	1	US-08-246-982A-5	Sequence 5, Appl
35	7.5	10366	1	US-08-453-265-5	Sequence 5, Appl
32	7.5	3819	1	US-07-686-322A-3	Sequence 3, Appl
32	7.5	3819	1	US-08-002-999-3	Sequence 3, Appl
32	7.5	3819	2	US-08-531-927B-3	Sequence 3, Appl
32	7.4	397	3	US-09-253-691-3	Sequence 3, Appl
35	7.4	477	3	US-09-135-994-1	Sequence 1, Appl
37	7.4	477	3	US-09-684-843A-1	Sequence 1, Appl
34.4	7.4	1332	2	US-08-481-814A-3	Sequence 3, Appl
34.4	7.4	1489	3	US-08-826-582-1	Sequence 81, App
34.4	7.4	1489	4	US-09-242-737-3	Sequence 1, Appl
34.4	7.4	1489	4	US-09-023-655-871	Sequence 12, Appl
34.4	7.4	1776	2	US-08-531-927B-1	Sequence 1, Appl
34.4	7.4	1776	3	US-09-041-886-12	Sequence 1, Appl
34.4	7.4	6177	4	US-08-479-913E-1	Sequence 1, Appl
34.4	7.4	7257	4	US-09-091-042A-1	Sequence 4, Appl
34.4	7.4	50000	4	US-09-146-053-4	Sequence 1, Appl
34.4	7.4	1664976	4	US-08-916-421B-1	Sequence 1, Appl
34.2	7.3	580073	4	US-08-545-528D-1	Sequence 3, Appl
34	7.3	300	3	US-09-135-994-3	Sequence 3, Appl
34	7.3	300	4	US-09-684-843A-3	Sequence 150, App
34	7.3	510	2	US-08-184-009-150	Sequence 150, App
34	7.3	510	2	US-08-458-356-150	Sequence 150, App
34	7.3	510	3	US-08-460-736-150	Sequence 150, App
34	7.3	510	4	US-09-535-370-150	Sequence 2, Appl
34	7.3	1472	4	US-09-634-368-2	Sequence 11, Appl
34	7.3	2013	4	US-09-634-368-11	Sequence 11, Appl
34	7.3	2060	1	US-08-480-547A-11	Sequence 11, Appl
34	7.3	2060	1	US-08-250-847B-11	Sequence 11, Appl
34	7.3	2060	3	US-08-463-949A-11	Sequence 11, Appl
34	7.3	2060	3	US-08-464-410A-11	Sequence 11, Appl
34	7.3	2060	5	PCT-US94-06066-11	Sequence 22, Appl
34	7.3	2645	1	US-08-480-547A-22	Sequence 22, Appl
34	7.3	2645	1	US-08-250-847B-22	Sequence 22, Appl
34	7.3	2645	2	US-08-463-949A-22	Sequence 22, Appl
34	7.3	2645	3	US-08-464-410A-22	Sequence 22, Appl
34	7.3	2645	5	PCT-US94-06066-22	Sequence 22, Appl
34	7.3	2709	4	US-09-634-368-13	Sequence 13, Appl
34	7.2	3215	4	US-08-948-113D-24	Sequence 24, Appl
34	7.2	3376	1	US-08-320-559-29	Sequence 29, Appl
34	7.2	3376	3	US-08-545-860D-29	Sequence 29, Appl
34	7.2	3376	5	PCT-US94-04496-29	Sequence 29, Appl
34	7.2	168	1	US-08-469-802B-4	Sequence 4, Appl
34	7.2	168	2	US-08-267-803B-4	Sequence 4, Appl
34	7.2	171	1	US-08-469-802B-5	Sequence 5, Appl
34	7.2	171	1	US-08-267-803B-5	Sequence 5, Appl
34	7.2	195	1	US-08-469-802B-2	Sequence 2, Appl
34	7.2	195	2	US-08-267-803B-2	Sequence 2, Appl
34	7.2	234	1	US-08-469-802B-3	Sequence 3, Appl
34	7.2	234	2	US-08-267-803B-3	Sequence 3, Appl
34	7.2	1037	4	US-09-181-585-3	Sequence 3, Appl
34	7.2	1159	4	US-09-181-585-1	Sequence 2, Appl
34	7.2	1471	4	US-09-181-585-2	Sequence 5, Appl
34	7.2	2892	1	US-08-264-534-5	Sequence 5, Appl
34	7.2	2892	1	US-08-083-590A-1	Sequence 5, Appl
34	7.2	2892	1	US-08-465-500-5	Sequence 5, Appl
34	7.2	2892	2	US-08-346-128-5	Sequence 5, Appl
34	7.2	2892	3	US-08-532-384-1	Sequence 5, Appl
34	7.2	2892	3	US-08-893-828-5	Sequence 7, Appl
34	7.2	2950	5	PCT-US93-08386-7	Sequence 1, Appl
34	7.2	2968	5	PCT-US93-08386-1	Sequence 1, Appl
34	7.2	3366	2	US-08-469-802B-1	Sequence 1, Appl
34	7.2	3366	2	US-08-267-803B-1	Sequence 14, Appl
34	7.2	7218	1	US-08-232-463-14	Sequence 8, Appl
34	7.2	10660	3	US-08-267-803B-8	Sequence 16, Appl
34	7.2	10660	3	US-09-041-886-16	Sequence 8, Appl
34	7.2	270	4	US-09-146-054-8	Sequence 8, Appl
34	7.2	270	4	US-09-664-977A-8	Sequence 7, Appl
34	7.2	506	1	US-08-469-802B-7	Sequence 7, Appl

```

; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match      8.7%; Score 40.4; DB 1; Length 7218;
Best Local Similarity 2.3%; Pred. No. 0.034;
Matches 5; Conservative 134; Mismatches 75; Indels 0; Gaps 0;

QY      6 TTGCGCTATCCACCTCCCCCAAGCGCTCATGCTGGAGTGGGACTGGTGGTGGCCAGAAAGTCTC 65
Db      1229 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1288

QY      66 GCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCCAGAAAGTCTC 125
Db      1289 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1348

QY      126 TTGCGCACTGAGCCGCCATCAGGATGGGCTTCTTCCGCCCTTCTTCTTCTGCT 185
Db      1349 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1408

QY      186 CTGCGCTCATGCGCTGCCATGACCTGACGCA 219
Db      1409 YYYYYYYYYYYYYYYYYYYYYYYYYTACCATA 1442

RESULT 3
US-09-521-511C-10/c
; Sequence 10, Application US/09521511C
; Patent No. 655358
; GENERAL INFORMATION:
; APPLICANT: Gurnett, Anne
; APPLICANT: Liberator, Paul A.
; APPLICANT: Donald, Robert
; APPLICANT: Schmatz, Dennis
; APPLICANT: Harris, Georgianna
; APPLICANT: Rattray, Sandra J.
; TITLE OF INVENTION: CYCLIC GMP DEPENDENT PROTEIN KINASE AS A

ALIGNMENTS

RESULT 1
US-09-146-053-4/c
; Sequence 4, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.
; TITLE OF INVENTION: Human Aminopeptidase P Gene
; FILE REFERENCE: MG103
; CURRENT APPLICATION NUMBER: US/09/146,053A
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/057,854
; EARLIER FILING DATE: 1997-09-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-146-053-4

Query Match      100.0%; Score 467; DB 4; Length 50000;
Best Local Similarity 100.0%; Pred. No. 1.5e-136;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CAATGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTG 60
Db      1518 CAATGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTG 1459

QY      61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGTCGCCAGAAA 120
Db      1458 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGTCGTCGCCAGAAA 1399

QY      121 GTCTCTTCTGCACTGAGGCCGCCATCAGGATGGGCTTCTTCCCTTCTGCTGCTG 180
Db      1398 GTCTCTTCTGCACTGAGGCCGCCATCAGGATGGGCTTCTTCCCTTCTGCTGCTG 1339

QY      181 TGCTCTGCTGCTCATGCGCTGCCATGACCTGCAAGCAAGCCAGCCCGCTGGGAGGG 240
Db      1338 TGCTCTGCTGCTCATGCGCTGCCATGACCTGCAAGCAAGCCCGCTGGGAGGG 1279

QY      241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 300
Db      1278 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTG 1219

QY      301 GGCTAGGGGGCTGCTTATTAAGTGGTGTGTTATGATCTTATCTATTAATTAACAA 360
Db      1218 GGCTAGGGGGCTGCTTATTAAGTGGTGTGTTATGATCTTATCTATTAATTAACAA 1159

QY      361 AGATATTAAAGCCCTGTTCATTAGAAATTTCTCCCTTCCCTGTGTTCAATGTTGTAA 420
Db      1158 AGATATTAAAGCCCTGTTCATTAGAAATTTCTCCCTTCCCTGTGTTCAATGTTGTAA 1099

QY      421 AGATGTTCTGTGTAATATGCTTTTATAATAACAGTTAAAGCTG 467
Db      1098 AGATGTTCTGTGTAATATGCTTTTATAATAACAGTTAAAGCTG 1052

RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

```

[illegible]

```

RESULT 6
US-09-115-446-5/c
; Sequence 5, Application US/09115446
; Patent No. 6165719
; GENERAL INFORMATION:
; APPLICANT: Chandy, George K.
; APPLICANT: Gargus, Jay J.
; APPLICANT: Gutman, George
; APPLICANT: Fantino, Emmanuelle
; APPLICANT: Kalman, Katarin
; TITLE OF INVENTION: hKCA3/KCNN3 SMALL CONDUCTANCE CALCIUM
; TITLE OF INVENTION: ACTIVATED POTASSIUM CHANNEL: A DIAGNOSTIC
; TITLE OF INVENTION: MARKER AND THERAPEUTIC TARGET
; FILE REFERENCE: 07306/014001
; CURRENT APPLICATION NUMBER: US/09/115.446
; CURRENT FILING DATE: 1998-07-14
; EARLIER APPLICATION NUMBER: 60/052,556
; EARLIER FILING DATE: 1997-07-15
; EARLIER APPLICATION NUMBER: 60/070,741
; EARLIER FILING DATE: 1998-01-08
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2526
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-115-446-5

```

Query Match



```

; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5670362el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,853
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; CLONE: Bam20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-853-7

```

```

Query Match      7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      6 TTTCCTATCCACTCCCAAGCCCTTACCTATGCTGCTAAAGCTGCTGCT 65
Db      461 TTTTCGCTCAACTGAGCTGCTGAGCTGCACTGCTGCTGCTGCTGCT 402

QY      66 GCTGCTGCTGCTTAAAGCTCATGCTT 92
Db      401 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375

```

```

RESULT 10
US-08-468-855-7/c
; Sequence 7, Application US/08468855
; Patent No. 5780289
; GENERAL INFORMATION:
; APPLICANT: van den BOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE

```

```

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5780289el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,855
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; CLONE: Bam20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-855-7

```

```

Query Match      7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      6 TTTCCTATCCACTCCCAAGCCCTTACCTATGCTGCTGCTAAAGCTGCTGCT 65
Db      461 TTTTCGCTCAACTGAGCTGCTGAGCTGCACTGCTGCTGCTGCTGCT 402

QY      66 GCTGCTGCTGCTTAAAGCTCATGCTT 92
Db      401 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375

```

```

RESULT 11
US-08-310-357-7/c
; Sequence 7, Application US/08310357
; Patent No. 5789233
; GENERAL INFORMATION:
; APPLICANT: van den BOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Organon Teknika Corporation
; ADDRESSEE: Biotechnology Research Institute
; STREET: 1330-A Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310,357
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/904,075
; FILING DATE:
; FILING DATE: 18-Jun-1991
; APPLICATION NUMBER: EP 91.201.523.7
; ATTORNEY/AGENT INFORMATION:
; NAME: William M. Blackstone
; REGISTRATION NUMBER: 29,772
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; IMMEDIATE SOURCE:
; CLONE: Em20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-310-357-7

```

```

Query Match          7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      6 TTGGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCT 65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      461 TTTTGTCTCCAACTGCGAGTGTGCGAGTGTGCGAACTGCTGCTGCTGCTGCTGCT 402

QY      66 GTGCTGCTGCTTAAAGGCTCATGCTT 92
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      401 GCTGCTGCTGCTGCCATGGTTCTTCAT 375

```

```

RESULT 12
US-08-468-852-7/c
; Sequence 7, Application US/08468852
; Patent No. 5792644
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5792644el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,852
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-Jun-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gornley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; IMMEDIATE SOURCE:
; CLONE: Em20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-852-7

```

```

Query Match          7.7%; Score 35.8; DB 1; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24;
Matches 55; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY      6 TTGGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCT 65
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      461 TTTTGTCTCCAACTGCGAGTGTGCGAGTGTGCGAACTGCTGCTGCTGCTGCTGCT 402

QY      66 GTGCTGCTGCTTAAAGGCTCATGCTT 92
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      401 GCTGCTGCTGCTGCCATGGTTCTTCAT 375

```

```

RESULT 13
US-08-468-857-7/c
; Sequence 7, Application US/08468857
; Patent No. 5925347
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5925347el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,857
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,357
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,865
; FILING DATE: 06-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,075
; FILING DATE: 18-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91.201.523.7
; FILING DATE: 18-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34,409
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; IMMEDIATE SOURCE:
; CLONE: Bam20E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..508
; US-08-468-857-7
```

```
Query Match 7.7%; Score 35.8; DB 2; Length 800;
Best Local Similarity 63.2%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 32;

QY 6 TTTCCTATCCACTCCCAAGCCCTTTACCTATGCTGCTCTAAAGCTGCTGCT 65
Db 461 TTTCCTATCCACTCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 402

QY 66 GCTGCTGCTGCTTAAAGGCTCATGCTT 92
Db 401 GCTGCTGCTGCTGCCATGGTTCTTCAT 375
```

```
RESULT 14
US-09-205-995-75/c
; Sequence 75, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
; PRIOR APPLICATION NUMBER: 08/661,627
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 107
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
```

```
; OTHER INFORMATION: Description of Artificial Sequence: Reverse gene
; OTHER INFORMATION: construct corresponding to a specific region of
; OTHER INFORMATION: the mouse Ii gene.
US-09-205-995-75
```

```
Query Match 7.6%; Score 35.6; DB 4; Length 107;
Best Local Similarity 71.2%; Pred. No. 0.079; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 19;

QY 20 TCCCCAAGCCCTTTACCTATGCTGCTCTAAAGCTGCTGCTGCTGCTTAA 79
Db 73 TCCACAGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14

QY 80 AAGGCT 85
Db 13 GTGTCT 8
```

```
RESULT 15
US-09-205-995-72/c
; Sequence 72, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
; PRIOR APPLICATION NUMBER: 08/661,627
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse gene
; OTHER INFORMATION: construct corresponding to a specific region of
; OTHER INFORMATION: the mouse Ii gene.
US-09-205-995-72
```

```
Query Match 7.6%; Score 35.6; DB 4; Length 169;
Best Local Similarity 71.2%; Pred. No. 0.11; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 19;

QY 20 TCCCCAAGCCCTTTACCTATGCTGCTCTAAAGCTGCTGCTGCTGCTTAA 79
Db 73 TCCACAGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14

QY 80 AAGGCT 85
Db 13 GTGTCT 8
```

```
RESULT 16
US-09-205-995-77/c
; Sequence 77, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
```



[illegible]



```
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 306
; LENGTH: 9053
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 898877.6
; NAME/KEY: unsure
; LOCATION: 2006, 2012
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-306

Query Match          7.5%; Score 35.2; DB 4; Length 9053;
Best Local Similarity 54.9%; Pred. No. 1.7;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      8  TGCTATTCACCTCCCGGAGCCCTTTACCTATGCTGCTGCTAAAGCTGCTGCTGCTG 67
Db      2090  TTTCTATCTTTCTCCCTCCCTGGCTGCTGCGGGGATGGGCTGCTGCTGCTGCTGCTG 2031

QY      68  TGCTGCTGCTTAAAGCTCATGCTTGAGTGGGAGCTGCTGCTGCTGCTGCTGCTGCT 127
Db      2030  TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1971

QY      128  CT 129
Db      1970  CT 1969

RESULT 24
US-08-206-176-7
; Sequence 7, Application US/08206176
; Patent No. 5639940
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Dalrymple, Michael A
; APPLICANT: Prunkard, Donna E
; APPLICANT: Foster, Donald C
; TITLE OF INVENTION: Production of Fibrinogen in Transgenic
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 4225 Roosevelt Way, N.E.
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,176
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 93-15
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-547-8080 ext 322
; TELEFAX: 206-548-2329
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ovine beta-lactoglobulin
US-08-206-176-7

Query Match          7.5%; Score 35.2; DB 1; Length 10807;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY      129  TGCCACTGAGCGCCCATCAGGGATTGGCCCTTTCTTCCCTTCTTCTTCTTCTTCTTCT 188
Db      8265  TGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 8324

QY      189  GCCTCATCGGCTGCCATGACCTGCAGCCCAAGCCCGCCGAGGGAGGAGGAGGAGGAGGAG 248
Db      8325  GGGTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8384

QY      249  GGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGAGAGGAGGAGGAGGAGGAGGAG 308
Db      8385  GGGGCTGGGAGCCCAACATCAGGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8444

QY      309  GGGC 312
Db      8445  GCTC 8448

RESULT 25
US-08-756-506-5
; Sequence 5, Application US/08756506
; Patent No. 5905185
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,506
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 95-28
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 10807 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-756-506-5

Query Match          7.5%; Score 35.2; DB 2; Length 10807;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 91; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 129 TGCACCTGACGCCCCCATCAGGATTGGGCTTCTTTCCCTTCTCTTCTGTCTCCT 188
Db 8265 TGACTCTCCCTCCCCACAGGCGAGTGCACGCTAGGTGAGCCCTGCGCGTGCCTG 8324

QY 189 GCCTCATGGCTCGCATGACCTGCGAGCCAGCCAGCCCGCTGGGAGGGGAGAACT 248
Db 8325 GGGTAAGCTCGCTGCCCTGCCACGCTCTCTGGGCACACATCGGGGTAGGGGTCTTGGT 8384

QY 249 GGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGCTAGTGGGTGGCTAGGG 308
Db 8385 GGGGCTGGGACCCACATCAGGCCCTGGGGTCCCTCTGAGATGGCTGGAAGCTGG 8444

QY 309 GGGC 312
Db 8445 GGTG 8448

RESULT 26
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 17:
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match          7.5%; Score 35.2; DB 3; Length 176373;
Best Local Similarity 68.1%; Pred. No. 11;
Matches 49; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 7 TTGCTATCCACCTCCCCAGCCCTTACTATGCTGTGCTAAGCTGCTGCTGCTG 66
Db 95731 TGGACTCACACGAGTCCACTCCTGCTGTTCCTCTGCTTGTCTGAGGCTTCTGCTG 95790

QY 67 CTGCTGTGCTT 78
Db 95791 CTGCTGTGCTT 95802

RESULT 27
US-08-457-273B-41/c
; Sequence 41, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
```

```
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESS: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10348 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-457-273B-41

Query Match          7.5%; Score 35; DB 2; Length 10348;
Best Local Similarity 88.4%; Pred. No. 2.2;
Matches 38; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 41 TGCTGCTGCTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTAAAG 83
Db 404 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGGAAG 362

RESULT 28
US-08-556-419-13/c
; Sequence 13, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony
; APPLICANT: Worley, Paul
; APPLICANT: Snyder, Solomon
; TITLE OF INVENTION: Huntingtin-associated protein
; FILE REFERENCE: 01107.52271
; CURRENT APPLICATION NUMBER: US/08/556,419C
; CURRENT FILING DATE: 1995-11-09
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Fast-Seq for Windows Version 3.0
; SEQ ID NO 13:
; LENGTH: 10348
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-556-419-13

Query Match          7.5%; Score 35; DB 3; Length 10348;
Best Local Similarity 88.4%; Pred. No. 2.2;
Matches 38; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

7 GENERAL INFORMATION:  
7 APPLICANT: MacDonald, Marcy E.  
7 APPLICANT: Ambrose, Christine M.  
7 APPLICANT: Duyao, Mabel P.  
7 APPLICANT: Gueella, James F.  
7 TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof  
7 NUMBER OF SEQUENCES: 25  
7 CORRESPONDENCE ADDRESS:  
7 ADDRESSEE: Sterne, Kessler, Goldstein & Fox

ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/453,265  
 FILING DATE: 30-MAY-1995  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ludwig, Steven R.  
 REGISTRATION NUMBER: 36,203  
 REFERENCE/DOCKET NUMBER: 0609.3980003  
 TELECOMMUNICATION INFORMATION:





```

Db      1027 TTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984
RESULT 39
US-08-836-582-1/c
; Sequence 1, Application US/08836582
; Patent No. 6045999
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; APPLICANT: Beijersbergen, Roderick L
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye
; STREET: 1100 No. 6045999th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,582
; FILING DATE: 14-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00868
; FILING DATE: 18-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
US-08-836-582-1
Query Match          7.4%; Score 34.4; DB 3; Length 1
Best Local Similarity 86.4%; Pred.No.1;
Matches 38; Conservative 0; Mismatches 6; Indels

Oy      34   TTACTATGCTGCTGCTGAAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 77
           ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      991 TTACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984

RESULT 40
US-09-265-566-1/c
; Sequence 1, Application US/09265566
; Patent No. 6303335
; GENERAL INFORMATION:
; APPLICANT: Bernards, Rene
; APPLICANT: Beijersbergen, Roderick L
; TITLE OF INVENTION: Transcription factor E2F-4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderbilt
; STREET: 1100 No. 6303335th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US

```



```

;
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,566
; FILING DATE: 10-Mar-1999
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,582
; FILING DATE: <Unknown>
; APPLICATION NUMBER: GB 9423049.7
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Arthur R. Crawford
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 620-18
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-265-566-1

Query Match 7.4%; Score 34.4; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 1;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 34 TTACTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTGCT 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 991 TTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948

RESULT 41
US-09-242-737-3/c
; Sequence 3, Application US/09242737
; Patent No. 6368809
; GENERAL INFORMATION:
; APPLICANT: BERNARDS, REN(
; TITLE OF INVENTION: E2F UBIQUITINATION DOMAIN, AND ASSAYS FOR
; INHIBITORS OF E2F UBIQUITINATION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,737
; FILING DATE: 23-Feb-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/02293
; FILING DATE: 22-AUG-1997
; APPLICATION NUMBER: GB 9617697.9
; FILING DATE: 23-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.

```

```

;
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-63
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 27..1268
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-242-737-3

Query Match 7.4%; Score 34.4; DB 4; Length 1489;
Best Local Similarity 86.4%; Pred. No. 1;
Matches 38; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 34 TTACTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTGCT 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 991 TTACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948

RESULT 42
US-09-023-655-871/c
; Sequence 871, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 871:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1061145

```







Db 389857 TTTTAACTGAGGTATATGACATAAATCAATCAATGAATGAATTTTAATGCAAT 389798  
QY 378 TCATTAAAGAAATGTTCCCTTCCCTGTTCAATGTTTGTAAAGATTGTTCTGTGTA 437  
Db 389797 ACTGTATATAATATATATCTGCTCTCAATCTCAATTAATTTATTTATTTGTTGGAAA 389738  
QY 438 TATG 441  
Db 389737 TATG 389734

## RESULT 49

US-08-545-528D-1/c  
; Sequence 1, Application US/08545528D  
; Patent No. 6537773  
; GENERAL INFORMATION:  
; APPLICANT: Fraser et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment  
; Patent No. 6537773  
; TITLE OF INVENTION: Thereof, and Uses Thereof  
; FILE REFERENCE: PB193PI  
; CURRENT APPLICATION NUMBER: US/08/545,528D  
; CURRENT FILING DATE: 1995-10-19  
; PRIOR APPLICATION NUMBER: US 08/488,018  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/473,545  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 580073  
; TYPE: DNA  
; ORGANISM: Mycoplasma genitalium  
US-08-545-528D-1

Query Match 7.3%; Score 34.2; DB 4; Length 580073;  
Best Local Similarity 52.4%; Pred. No. 49;  
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 317 TTATTTAACTGTTGTTTAAGTATCTATCTACTAATTTATACAAAGATATTAAAGGCCCTG 376  
Db 377708 TTGTGAGAGTGGTTATTATTGTTTGTGTTGTTGTTTTCAAATTTGTAACCCCTT 377649  
QY 377 TCATTAAAGAAATGTTCCCTTCCCTGTTCAATGTTTGTAAAGATTGTTCTGTGTA 436  
Db 377648 ATCAGTAGGACAGATCCATTGCTGTGTTGTTGTTTCAATTAATAAGGGGATGAAAAACA 377589  
QY 437 ATATGCTTTTATATAAACAGTT 459  
Db 377588 AACTGCAGTTTAAATTCACCT 377566

## RESULT 50

US-09-135-994-3/c  
; Sequence 3, Application US/09135994A  
; Patent No. 6280938  
; GENERAL INFORMATION:  
; APPLICANT: Randum et al.  
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE  
; FILE REFERENCE: University of Minnesota  
; CURRENT APPLICATION NUMBER: US/09/135,994A  
; CURRENT FILING DATE: 1998-08-18  
; EARLIER APPLICATION NUMBER: 60/056,170  
; EARLIER FILING DATE: 1997-08-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 300  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-135-994-3  
Query Match 7.3%; Score 34; DB 3; Length 300;

Best Local Similarity 80.0%; Pred. No. 0.49;  
Matches 40; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
QY 28 GCCCCTTTACCTATGCTGCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 77  
Db 193 GGCCTCTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 144  
Search completed: May 25, 2004, 17:19:51  
Job time : 101 secs

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 17:19:57 ; Search time 92 Seconds  
(without alignments)  
2816.979 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctaccact.....taataaacagttaaagctg 467

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 10

Total number of hits satisfying chosen parameters: 144238

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : Issued Patents.NA.\*

- 1: /cgm2\_6/ptodata/2/ina/5A.COMB.seq.\*
- 2: /cgm2\_6/ptodata/2/ina/5B.COMB.seq.\*
- 3: /cgm2\_6/ptodata/2/ina/6A.COMB.seq.\*
- 4: /cgm2\_6/ptodata/2/ina/6B.COMB.seq.\*
- 5: /cgm2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*
- 6: /cgm2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	467	100.0	50000	4	US-09-146-053-4
2	25	5.4	36	2	US-08-863-639A-31
3	25	5.4	91	4	US-09-091-952A-15
4	25	5.4	94	4	US-09-091-952A-14
5	25	5.4	397	3	US-09-253-691-3
6	25	5.4	432	4	US-09-060-299-5
7	25	5.4	432	4	US-09-402-923A-5
8	25	5.4	570	4	US-09-621-976-91
9	25	5.4	575	4	US-09-621-976-2968
10	25	5.4	841	4	US-09-566-921-136
11	25	5.4	1776	2	US-08-531-927B-1
12	25	5.4	1776	3	US-09-041-886-12
13	25	5.4	1884	3	US-08-753-007A-5
14	25	5.4	1884	3	US-09-398-496-5
15	25	5.4	2023	4	US-09-491-522-6
16	25	5.4	2128	4	US-09-675-305-13
17	25	5.4	2234	1	US-08-261-822A-7
18	25	5.4	2234	5	PCT-US95-07744A-7
19	25	5.4	2450	4	US-09-491-522-2
20	25	5.4	2770	3	US-09-008-697A-13
21	25	5.4	3231	4	US-09-066-028-14
22	25	5.4	4262	4	US-09-521-511C-10
23	25	5.4	4843	4	US-09-060-299-2
24	25	5.4	4843	4	US-09-402-923A-2
25	25	5.4	5098	4	US-09-060-299-1
26	25	5.4	5098	4	US-09-402-923A-1
27	25	5.4	6692	4	US-09-491-522-1

8065	4	US-09-091-952A-6	Sequence 6, Appli
5.4	1	US-08-068-747-6	Sequence 6, Appli
5.1	30	US-08-068-747-11	Sequence 11, Appl
5.1	30	US-08-068-747-11	Sequence 30, Appl
5.1	30	US-08-863-639A-30	Sequence 4, Appli
5.1	30	US-09-135-994-4	Sequence 4, Appli
5.1	30	US-09-684-843A-4	Sequence 14, Appl
5.1	31	US-08-570-155-14	Sequence 14, Appl
5.1	31	PCT-US95-02861-14	Sequence 29, Appl
5.1	33	US-08-863-639A-29	Sequence 1, Appli
5.1	51	US-08-068-747-1	Sequence 50, Appl
5.1	53	US-09-136-080E-50	Sequence 14, Appl
5.1	57	US-09-043-303-14	Sequence 13, Appl
5.1	69	US-09-043-303-13	Sequence 10, Appl
5.1	75	US-09-043-303-10	Sequence 11, Appl
5.1	78	US-09-043-303-11	Sequence 12, Appl
5.1	78	US-09-043-303-12	Sequence 75, Appl
5.1	107	US-09-205-995-75	Sequence 6, Appli
5.1	154	US-08-469-802B-6	Sequence 6, Appli
5.1	154	US-08-267-803B-6	Sequence 13, Appl
5.1	160	US-08-332-766A-13	Sequence 17, Appl
5.1	165	US-08-043-303-17	Sequence 4, Appli
5.1	168	US-08-469-802B-4	Sequence 4, Appli
5.1	168	US-08-267-803B-4	Sequence 72, Appl
5.1	169	US-09-205-995-72	Sequence 5, Appli
5.1	171	US-08-469-802B-5	Sequence 5, Appli
5.1	171	US-08-267-803B-5	Sequence 16, Appl
5.1	183	US-09-707-919A-16	Sequence 77, Appl
5.1	190	US-08-205-995-77	Sequence 2, Appli
5.1	195	US-08-469-802B-2	Sequence 2, Appli
5.1	195	US-08-267-803B-2	Sequence 7, Appli
5.1	203	US-09-043-303-7	Sequence 13, Appl
5.1	228	5273901-10	Patent No. 5273901
5.1	228	5482709-9	Sequence 3, Appli
5.1	234	US-08-469-802B-3	Sequence 3, Appli
5.1	234	US-08-267-803B-3	Sequence 8, Appli
5.1	270	US-09-146-054-8	Sequence 8, Appli
5.1	270	US-09-664-977A-8	Sequence 21, Appl
5.1	283	US-08-332-766A-21	Sequence 6, Appli
5.1	286	US-08-332-766A-6	Sequence 3, Appli
5.1	300	US-09-135-994-3	Sequence 3, Appli
5.1	300	US-09-684-843A-3	Sequence 3, Appli
5.1	325	US-08-531-927B-3	Sequence 5, Appli
5.1	412	US-09-612-033B-5	Sequence 17616, A
5.1	458	US-09-621-976-17616	Sequence 13, Appl
5.1	459	US-09-707-919A-13	Sequence 1, Appli
5.1	477	US-09-135-994-1	Sequence 1, Appli
5.1	477	US-09-684-843A-1	Sequence 13, Appl
5.1	490	US-09-679-686B-13	Sequence 7, Appli
5.1	506	US-08-469-802B-7	Sequence 7, Appli
5.1	506	US-08-267-803B-7	Sequence 150, App
5.1	510	US-08-184-009-150	Sequence 150, App
5.1	510	US-08-458-356-150	Sequence 150, App
5.1	510	US-08-460-736-150	Sequence 150, App
5.1	510	US-09-535-370-150	Sequence 1, Appli
5.1	516	US-09-083-268-1	Sequence 5, Appli
5.1	623	US-09-043-303-5	Sequence 915, App
5.1	688	US-08-998-416-915	Sequence 9, Appli
5.1	702	US-09-612-033B-9	Sequence 590, App
5.1	715	US-08-998-416-590	Sequence 3, Appli
5.1	1037	US-09-181-585-3	Sequence 2, Appli
5.1	1080	US-08-918-727-2	Sequence 2, Appli
5.1	1080	US-09-205-680A-2	Sequence 1, Appli
5.1	1093	US-09-010-147B-1	Sequence 1, Appli
5.1	1159	US-09-181-585-1	Sequence 13, Appl
5.1	1200	US-09-612-033B-13	Sequence 3, Appli
5.1	1332	US-08-481-814A-3	Sequence 1, Appli
5.1	1383	US-08-484-044-1	Sequence 2, Appli
5.1	1471	US-09-181-585-2	Sequence 2, Appli
5.1	1472	US-09-634-368-2	Sequence 1, Appli
5.1	1489	US-08-836-582-1	Sequence 1, Appli
5.1	1489	US-09-265-566-1	Sequence 3, Appli
5.1	1489	US-09-242-737-3	Sequence 871, App
5.1	1489	US-09-023-655-871	

## ALIGNMENTS

RESULT 1  
 US-09-146-053-4/c  
 ; Sequence 4, Application US/09146053A  
 ; Patent No. 6399349  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ryan, James W.  
 ; APPLICANT: Sprinkle, Terry Joe Curtis  
 ; APPLICANT: Venema, Richard C.  
 ; TITLE OF INVENTION: Human Aminopeptidase P Gene  
 ; FILE REFERENCE: MCG103  
 ; CURRENT APPLICATION NUMBER: US/09/146,053A  
 ; CURRENT FILING DATE: 1998-09-02  
 ; EARLIER APPLICATION NUMBER: 60/057,854  
 ; EARLIER FILING DATE: 1997-09-02  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 50000  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-146-053-4

Query Match 100.0%; Score 467; DB 4; Length 50000;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-223;  
 Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CAATGTTTCCTATCCACTCCCGCCCAAGCCCTTTACCTATCTGCTGCTAAGCTGCTG 60
DB	1518	CAATGTTTCCTATCCACTCCCGCCCAAGCCCTTTACCTATCTGCTGCTAAGCTGCTG 1459
QY	61	CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTTGGAGTGGGGACTGGTGGTCCAGAAA 120
DB	1458	CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTTGGAGTGGGGACTGGTGGTCCAGAAA 1399
QY	121	GTCTCTTCTGCCACTGACGCCGCCCATCAGGAGTGGGCTTTTCCCGCTTCCTTTCTG 180
DB	1398	GTCTCTTCTGCCACTGACGCCGCCCATCAGGAGTGGGCTTTTCCCGCTTCCTTTCTG 1339
QY	181	TCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB	1338	TCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279
QY	241	GAGAAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAGAGGGTAGTGGGTG 300
DB	1278	GAGAAAGTGGGGATGGCTTAAGAAAGCTGGGAGATAGGAAACAGAGAGGGTAGTGGGTG 1219
QY	301	GGCTAGGGGGCTGCTTATTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360
DB	1218	GGCTAGGGGGCTGCTTATTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1159
QY	361	AGATATTAAAGGCTGCTTATTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 420
DB	1158	AGATATTAAAGGCTGCTTATTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1099
QY	421	AGATTGTTCTGTTAAATGCTTTTAAATGCTTTTAAATGCTTTTAAATGCTTTTAAATG 467
DB	1098	AGATTGTTCTGTTAAATGCTTTTAAATGCTTTTAAATGCTTTTAAATGCTTTTAAATG 1052

RESULT 2  
 US-08-863-639A-31/c  
 ; Sequence 31, Application US/08863639A  
 ; Patent No. 5981185  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Matson, Robert S.  
 ; APPLICANT: Coassin, Peter J.  
 ; APPLICANT: Rampal, Jang B.  
 ; APPLICANT: Caskey, C. T.

TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS  
 NUMBER OF SEQUENCES: 95  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sheldon & Mak  
 STREET: 225 South Lake Avenue, 9th Floor  
 CITY: Pasadena  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 91101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: Corel WordPerfect 8 version  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/863,639A  
 FILING DATE: May 28, 1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Joseph E. Mueth  
 REGISTRATION NUMBER: 20,532  
 REFERENCE/DOCKET NUMBER: 11859-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (626) 796-4000  
 TELEFAX: (626) 795-6321  
 INFORMATION FOR SEQ ID NO: 31:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 36 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: Other nucleic acid  
 US-08-863-639A-31

Query Match 5.4%; Score 25; DB 2; Length 36;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	53	CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 77
DB	27	CGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3

RESULT 3  
 US-09-091-952A-15  
 ; Sequence 15, Application US/09091952A  
 ; Patent No. 6458532  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Detera-Wadleigh, Sevilla D.  
 ; Gershon, Elliot S.  
 ; Radner, Judith A.  
 ; Goldin, Lynn R.  
 ; Berrettini, Wade H.  
 ; Yoshikawa, Takeo  
 ; Sanders, Alan R.  
 ; Esterling, Lisa E.  
 ; TITLE OF INVENTION: Chromosomal Markers and Diagnostic Tests for Manic-Depressive Illness  
 NUMBER OF SEQUENCES: 197  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Crew LLP  
 STREET: Two Embarcadero Center, Eighth Floor  
 CITY: San Francisco  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94111-3834  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/091,952A

```

; FILING DATE: 19-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,278
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: PCT/US97/19381
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-297100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 91 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1...91
; OTHER INFORMATION: Clone 22 allele 2 polymorphic marker
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: 37...63
; OTHER INFORMATION: Clone 22 allele 2 polymorphic repeat
; sequence
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-091-952A-15

Query Match 5.4%; Score 25; DB 4; Length 91;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 40 GCTGCTGCTGCTGCTGCTGCTGCTT 64

RESULT 4
US-09-091-952A-14
; Sequence 14, Application US/09091952A
; Patent No. 6458532
; GENERAL INFORMATION:
; APPLICANT: Detera-Wadleigh, Sevilla D.
; Gershon, Elliot S.
; Badher, Judith A.
; Goldin, Lynn R.
; Berrettini, Wade H.
; Yoshikawa, Takeo
; Sanders, Alan R.
; Esterling, Lisa E.
; TITLE OF INVENTION: Chromosomal Markers and Diagnostic
; Tests for Manic-Depressive Illness
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,952A

```

```

; FILING DATE: 19-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,278
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: PCT/US97/19381
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-297100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 94 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1...94
; OTHER INFORMATION: Clone 22 allele 1 polymorphic marker
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: 37...66
; OTHER INFORMATION: Clone 22 allele 1 polymorphic repeat
; sequence
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-091-952A-14

Query Match 5.4%; Score 25; DB 4; Length 94;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 43 GCTGCTGCTGCTGCTGCTGCTGCTT 67

RESULT 5
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 5.4%; Score 25; DB 3; Length 397;
Best Local Similarity 100.0%; Pred. No. 0.0049;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCTT 77
Db 357 CGCTGCTGCTGCTGCTGCTGCTGCTT 333

```



```
RESULT 6
US-09-060-299-5
; Sequence 5, Application US/09060299
; Patent No. 6545137
; GENERAL INFORMATION:
; APPLICANT: Todd, John A
; APPLICANT: Hess, John W
; APPLICANT: Caskey, Charles T
; APPLICANT: Cox, Roger D
; APPLICANT: Gerhold, David
; APPLICANT: Hammond, Holly
; APPLICANT: Hey, Patricia
; APPLICANT: Kawaguchi, Yoshihiko
; APPLICANT: Merriman, Tony R
; APPLICANT: Metzker, Michael L
; TITLE OF INVENTION: No. 6545137el Receptor
; NUMBER OF SEQUENCES: 455
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon and Vanderhye
; STREET: 1100 No. 6545137th Glebe Road, Eighth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,299
; FILING DATE: 15-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/043,553
; FILING DATE: 15-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,740
; FILING DATE: 05-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: B.J. Sadoff
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 620-35
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4091
; TELEFAX: (703)816-4100
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-060-299-5
; Query Match 5.4%; Score 25; DB 4; Length 432;
; Best Local Similarity 100.0%; Pred. No. 0.0049;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 80 CGCTGCTGCTGCTGCTGCTGCTGCT 104

RESULT 7
US-09-402-923A-5
; Sequence 5, Application US/09402923A
; Patent No. 6555654
; GENERAL INFORMATION:
; APPLICANT: Todd, John A
; APPLICANT: Hess, John W
; APPLICANT: Caskey, Charles T
; APPLICANT: Cox, Roger D
; APPLICANT: Gerhold, David
; APPLICANT: Hammond, Holly
; TITLE OF INVENTION: No. 6555654el LDL-Receptor
; NUMBER OF SEQUENCES: 455
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon and Vanderhye
; STREET: 1100 No. 6555654th Glebe Road, Eighth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/402,923A
; FILING DATE: 14-Feb-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB98/01102
; FILING DATE: 15-APR-1998
; APPLICATION NUMBER: US 60/043,553
; FILING DATE: 15-APR-1997
; APPLICATION NUMBER: US 60/048,740
; FILING DATE: 05-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: B.J. Sadoff
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 620-81
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4091
; TELEFAX: (703)816-4100
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-402-923A-5
; Query Match 5.4%; Score 25; DB 4; Length 432;
; Best Local Similarity 100.0%; Pred. No. 0.0049;
; Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 80 CGCTGCTGCTGCTGCTGCTGCTGCT 104

RESULT 8
US-09-621-976-91/c
; Sequence 91, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 91
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
```

```

; LOCATION: 355..570
; NAME/KEY: sig peptide
; LOCATION: 355..564
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 7.30000019073486
; OTHER INFORMATION: seq YFLSLPLFLFA/GA
US-09-621-976-91

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0049; Length 570;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 458 GCTGCTGCTGCTGCTGCTGCTGCTT 434

RESULT 9
US-09-621-976-2968/c
; Sequence 2968, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2968
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 116..286
US-09-621-976-2968

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0049; Length 575;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 458 GCTGCTGCTGCTGCTGCTGCTGCTT 434

RESULT 10
US-09-566-921-136
; Sequence 136, Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 136
; LENGTH: 841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6682888 993585.1
; NAME/KEY: unsure
; LOCATION: 610-611
; OTHER INFORMATION: a, t, c, g, or other

US-09-566-921-136
Query Match
Best Local Similarity 100.0%; Pred. No. 0.0051; Length 841;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTT 78
Db 583 GCTGCTGCTGCTGCTGCTGCTGCTT 607

RESULT 11
US-08-531-927B-1/c
; Sequence 1, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531,927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATH95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1115
US-08-531-927B-1

Query Match
Best Local Similarity 100.0%; Pred. No. 0.0051; Length 1776;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGTGTGCTGCTGCTGCTGCTGCTT 77
Db 988 CGTGTGCTGCTGCTGCTGCTGCTT 964

RESULT 12
US-09-041-886-12/c
; Sequence 12, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.

```

```

; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1116
; US-09-041-886-12

Query Match 5.4%; Score 25; DB 3; Length 1776;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCT 77
Db 988 CGCTGCTGCTGCTGCTGCTGCT 964

RESULT 13
US-08-753-007A-5/c
; Sequence 5, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,007A
; FILING DATE: 19-NOV-1996

```

```

; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1884 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 664...1883
; OTHER INFORMATION:
; US-08-753-007A-5

Query Match 5.4%; Score 25; DB 3; Length 1884;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCT 77
Db 65 CGCTGCTGCTGCTGCTGCTGCT 41

RESULT 14
US-09-398-496-5/c
; Sequence 5, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,496
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/753,007
; FILING DATE: 19-NOV-1996
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:

```



```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-261-822A-7

Query Match          5.4%; Score 25; DB 1; Length 2234;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCT 78
Db 1973 GCTGCTGCTGCTGCTGCTGCT 1949

RESULT 18
PCT-US95-07744A-7/c
; Sequence 7, Application PC/TUS9507744A
; GENERAL INFORMATION:
; APPLICANT: Trustees of The University of Pennsylvania
; TITLE OF INVENTION: Plant Genes for Sensitivity to Ethylene
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07744A
; FILING DATE: 15-JUNE-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/261,822
; FILING DATE: June 17, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-07744A-7

Query Match          5.4%; Score 25; DB 5; Length 2234;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCT 78
Db 1973 GCTGCTGCTGCTGCTGCTGCT 1949

RESULT 19
US-09-491-522-2
; Sequence 2, Application US/09491522
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2450 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-491-522-2

Query Match          5.4%; Score 25; DB 4; Length 2450;
Best Local Similarity 100.0%; Pred. No. 0.0051;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCT 77
Db 125 CGCTGCTGCTGCTGCTGCTGCT 149

RESULT 20
US-09-008-697A-13/c
; Sequence 13, Application US/09008697A
; Patent No. 6197504
; GENERAL INFORMATION:
; APPLICANT: Chow, King Lau
; TITLE OF INVENTION: USES OF MAB-21
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/09/008,697A  
FILING DATE: January 19, 1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Chan, Albert Wai-Kit  
REGISTRATION NUMBER: 36,479  
REFERENCE/DOCKET NUMBER: 50752-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2770 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-008-697A-13

Query Match 5.4%; Score 25; DB 3; Length 2770;  
Best Local Similarity 100.0%; Pred. No. 0.0051;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTT 78  
Db 619 GCTGCTGCTGCTGCTGCTGCTT 595

RESULT 21  
US-09-866-028-14/c  
Sequence 14, Application US/09866028  
Patent No. 6642360  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin  
APPLICANT: Botstein, David  
APPLICANT: Eaton, Dan  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Geritsen, Mary  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul  
APPLICANT: Grimaldi, Christopher  
APPLICANT: Gurney, Austin  
APPLICANT: Hillan, Kenneth  
APPLICANT: Kijavini, Ivar  
APPLICANT: Napier, Mary  
APPLICANT: Roy, Margaret  
APPLICANT: Tumas, Daniel  
APPLICANT: Wood, William  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P2548P1C1  
CURRENT APPLICATION NUMBER: US/09/866,028  
CURRENT FILING DATE: 2001-05-25  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 120  
SEQ ID NO 14  
LENGTH: 3231  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-09-866-028-14

Query Match 5.4%; Score 25; DB 4; Length 3231;  
Best Local Similarity 100.0%; Pred. No. 0.0051;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCT 77  
Db 459 CGCTGCTGCTGCTGCTGCTGCT 435

RESULT 22  
US-09-521-511C-10/c

Sequence 10, Application US/09521511C  
Patent No. 6555358  
GENERAL INFORMATION:  
APPLICANT: Gurnett, Anne  
APPLICANT: Liberator, Paul A.  
APPLICANT: Donald, Robert  
APPLICANT: Schmatz, Dennis  
APPLICANT: Harris, Georgianna  
APPLICANT: Rattray, Sandra J.  
TITLE OF INVENTION: CYCLIC GMP DEPENDENT PROTEIN KINASE AS A  
CHEMOTHERAPEUTIC TARGET FOR ANTIPROTOZOAL AGENTS  
FILE REFERENCE: 20362  
CURRENT APPLICATION NUMBER: US/09/521,511C  
CURRENT FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/129,058  
PRIOR FILING DATE: 1999-04-13  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 4262  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: PKG  
US-09-521-511C-10

Query Match 5.4%; Score 25; DB 4; Length 4262;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGTGTGCTGCTGCTGCTGCTGCT 77  
Db 2091 CGTGTGCTGCTGCTGCTGCTGCT 2067

RESULT 23  
US-09-060-299-2  
Sequence 2, Application US/09060299  
Patent No. 6545137  
GENERAL INFORMATION:  
APPLICANT: Todd, John A  
APPLICANT: Hess, John W  
APPLICANT: Caskey, Charles T  
APPLICANT: Cox, Roger D  
APPLICANT: Gerhold, David  
APPLICANT: Hammond, Holly  
APPLICANT: Hey, Patricia  
APPLICANT: Kawaguchi, Yoshihiko  
APPLICANT: Merriman, Tony R  
APPLICANT: Metzker, Michael L  
TITLE OF INVENTION: No. 6545137el Receptor  
NUMBER OF SEQUENCES: 455  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon and Vanderhye  
STREET: 1100 No. 6545137th Giebe Road, Eighth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: US  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/060,299  
FILING DATE: 15-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/043,553  
FILING DATE: 15-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/048,740  
FILING DATE: 05-JUN-1997

ATTORNEY/AGENT INFORMATION:  
NAME: B.J.Sadoff  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 620-35  
TELEPHONE: (703)816-4091  
TELEFAX: (703)816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4843 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-060-299-2

Query Match 5.4%; Score 25; DB 4; Length 4843;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCT 77  
DB 32 CGCTGCTGCTGCTGCTGCTGCT 56

RESULT 25  
US-09-060-299-1  
Sequence 1, Application US/09060299  
Patent No. 6545137  
GENERAL INFORMATION:  
APPLICANT: Todd, John A  
APPLICANT: Hess, John W  
APPLICANT: Caskey, Charles T  
APPLICANT: Cox, Roger D  
APPLICANT: Gerhold, David  
APPLICANT: Hammond, Holly  
APPLICANT: Hey, Patricia  
APPLICANT: Kawaguchi, Yoshihiko  
APPLICANT: Merriman, Tony R  
APPLICANT: Metzker, Michael L  
TITLE OF INVENTION: No. 6545137e1 Receptor  
NUMBER OF SEQUENCES: 455  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon and Vanderhye  
STREET: 1100 No. 6545137th Glebe Road, Eighth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: US  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION NUMBER: US/09/060,299  
FILING DATE: 15-APR-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/043,553  
FILING DATE: 15-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/048,740  
FILING DATE: 05-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: B.J.Sadoff  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 620-35  
TELEPHONE: (703)816-4091  
TELEFAX: (703)816-4100  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5098 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-060-299-1

Query Match 5.4%; Score 25; DB 4; Length 5098;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCT 77

ATTORNEY/AGENT INFORMATION:  
NAME: B.J.Sadoff  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 620-35  
TELEPHONE: (703)816-4091  
TELEFAX: (703)816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4843 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-060-299-2

Query Match 5.4%; Score 25; DB 4; Length 4843;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCT 77  
DB 32 CGCTGCTGCTGCTGCTGCTGCT 56

RESULT 24  
US-09-402-923A-2  
Sequence 2, Application US/09402923A  
Patent No. 6555654  
GENERAL INFORMATION:  
APPLICANT: Todd, John A  
APPLICANT: Hess, John W  
APPLICANT: Caskey, Charles T  
APPLICANT: Cox, Roger D  
APPLICANT: Gerhold, David  
APPLICANT: Hammond, Holly  
APPLICANT: Hey, Patricia  
APPLICANT: Kawaguchi, Yoshihiko  
APPLICANT: Merriman, Tony R  
APPLICANT: Metzker, Michael L  
TITLE OF INVENTION: No. 6555654e1 LDL-Receptor  
NUMBER OF SEQUENCES: 455  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon and Vanderhye  
STREET: 1100 No. 6555654th Glebe Road, Eighth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: US  
ZIP: VA 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION NUMBER: US/09/402,923A  
FILING DATE: 14-Feb-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB98/01102  
FILING DATE: 15-APR-1998  
APPLICATION NUMBER: US 60/043,553  
FILING DATE: 15-APR-1997  
APPLICATION NUMBER: US 60/048,740  
FILING DATE: 05-JUN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: B.J.Sadoff  
REGISTRATION NUMBER: 36,663  
REFERENCE/DOCKET NUMBER: 620-81  
TELEPHONE: (703)816-4091  
TELEFAX: (703)816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4843 base pairs

```
Db      80 CGCTGCTGCTGCTGCTGCTGCT 104
|||||
RESULT 26
US-09-402-923A-1
; Sequence 1, Application US/09402923A
; Patent No. 6555654
; GENERAL INFORMATION:
;   APPLICANT: Todd, John A
;   APPLICANT: Hess, John W
;   APPLICANT: Caskey, Charles T
;   APPLICANT: Cox, Roger D
;   APPLICANT: Gerhold, David
;   APPLICANT: Hammond, Holly
;   APPLICANT: Hey, Patricia
;   APPLICANT: Kawaguchi, Yoshihiko
;   APPLICANT: Merriman, Tony R
;   APPLICANT: Metzker, Michael L
; TITLE OF INVENTION: No. 6555654e1 LDL-Receptor
; NUMBER OF SEQUENCES: 455
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon and Vanderhye
; STREET: 1100 No. 6555654th Glebe Road, Eighth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: US
; ZIP: VA 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/402,923A
; FILING DATE: 14-Feb-2001
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB98/01102
; FILING DATE: 15-APR-1998
; APPLICATION NUMBER: US 60/043,553
; FILING DATE: 15-APR-1997
; APPLICATION NUMBER: US 60/048,740
; FILING DATE: 05-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: B.J.Sadoff
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 620-81
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)816-4091
; TELEFAX: (703)816-4100
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5098 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-402-923A-1
Query Match      5.4%; Score 25; DB 4; Length 5098;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 CGCTGCTGCTGCTGCTGCTGCT 77
Db      80 CGCTGCTGCTGCTGCTGCTGCT 104
|||||
RESULT 27
US-09-491-522-1
; Sequence 1, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
;   APPLICANT: Collige, Alain
;   APPLICANT: Lapiere, Charles M.
;   APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE.
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-1
Query Match      5.4%; Score 25; DB 4; Length 6692;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      53 CGCTGCTGCTGCTGCTGCTGCT 77
Db      125 CGCTGCTGCTGCTGCTGCTGCT 149
|||||
RESULT 28
US-09-091-952A-6
; Sequence 6, Application US/09091952A
; Patent No. 6458532
; GENERAL INFORMATION:
;   APPLICANT: Detera-Wadleigh, Sevilla D.
;   APPLICANT: Gershon, Elliot S.
;   APPLICANT: Badner, Judith A.
;   APPLICANT: Goldin, Lynn R.
;   APPLICANT: Berrettini, Wade H.
;   APPLICANT: Yoshikawa, Takeo
;   APPLICANT: Sanders, Alan R.
;   APPLICANT: Esterling, Lisa E.
; TITLE OF INVENTION: Chromosomal Markers and Diagnostic
;                               Tests for Manic-Depressive Illness
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
```



```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/091/952A
FILING DATE: 19-Apr-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,278
FILING DATE: 28-OCT-1996
APPLICATION NUMBER: PCT/US97/19381
FILING DATE: 28-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Timothy L.
REGISTRATION NUMBER: 35,367
REFERENCE/DOCKET NUMBER: 015280-297100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 8065 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: -
LOCATION: 1...8065
OTHER INFORMATION: Clone 22
FEATURE:
NAME/KEY: CDS
LOCATION: 116...1036
OTHER INFORMATION: Clone 22 coding region
FEATURE:
NAME/KEY: misc feature
LOCATION: 452...505
OTHER INFORMATION: alternatively spliced portion
FEATURE:
NAME/KEY: misc feature
LOCATION: 5595...5685
OTHER INFORMATION: amplified region for genotyping
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-091-952A-6
Query Match 5.4%; Score 25; DB 4; Length 8065;
Best Local Similarity 100.0%; Pred. No. 0.0053;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTT 78
Db 5634 GCTGCTGCTGCTGCTGCTGCTT 5658

RESULT 29
US-068-747-6/c
Sequence 6, Application US/08068747
Patent No. 5695933
GENERAL INFORMATION:
APPLICANT: Schalling, Martin
APPLICANT: Hudson, Thomas J.
APPLICANT: Housman, David E.
TITLE OF INVENTION: Direct Determination of Expanded
REFERENCE/DOCKET NUMBER: MIT-6141
TELECOMMUNICATION INFORMATION:
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/068,747
FILING DATE: 28-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6141
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic"
US-08-068-747-6
Query Match 5.1%; Score 24; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCT 5

RESULT 30
US-08-068-747-11
Sequence 11, Application US/08068747
Patent No. 5695933
GENERAL INFORMATION:
APPLICANT: Schalling, Martin
APPLICANT: Hudson, Thomas J.
APPLICANT: Housman, David E.
TITLE OF INVENTION: Direct Determination of Expanded
REFERENCE/DOCKET NUMBER: MIT-6141
TELECOMMUNICATION INFORMATION:
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/068,747
FILING DATE: 28-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-6141
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic"
US-08-068-747-11

Query Match          5.1%; Score 24; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 3 GCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 31
US-08-863-639A-30/c
; Sequence 30, Application US/08863639A
; Patent No. 5981185
; GENERAL INFORMATION:
; APPLICANT: Matson, Robert S.
; APPLICANT: Coassin, Peter J.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Caskey, C. T.
; TITLE OF INVENTION: OLGONUCLEOTIDE REPEAT ARRAYS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Corel WordPerfect 8 version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863.639A
; FILING DATE: May 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph E. Mueh
; REGISTRATION NUMBER: 20,532
; REFERENCE/DOCKET NUMBER: 11859-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
US-08-863-639A-30

Query Match          5.1%; Score 24; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5

RESULT 32
US-09-135-994-4/c
; Sequence 4, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135.994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-4

Query Match          5.1%; Score 24; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5

RESULT 33
US-09-684-843A-4/c
; Sequence 4, Application US/09684843A
; Patent No. 6514755
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: Regents of the University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/684.843A
; CURRENT FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/056,170
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 09/135,994
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-684-843A-4

Query Match          5.1%; Score 24; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5

RESULT 34
US-08-570-155-14/c
; Sequence 14, Application US/08570155
; Patent No. 5962332
; GENERAL INFORMATION:
; APPLICANT: Singer, Robert H.
; APPLICANT: Taneja, Krishan L.
; TITLE OF INVENTION: DETECTION OF TRINUCLEOTIDE REPEATS
; TITLE OF INVENTION: BY IN SITU HYBRIDIZATION
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,155
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/399,499
; FILING DATE: 07 March 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/214,823
; FILING DATE: 17 March 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06353/011001
; TELEPHONE: (617) 542-8906
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-570-155-14
;
; Query Match 5.1%; Score 24; DB 2; Length 31;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5
;
; RESULT 35
; US-08-570-155-14/c
; Sequence 14, Application PC/TUS9502861
; GENERAL INFORMATION:
; APPLICANT: Singer, Robert H.
; APPLICANT: Taneja, Krishan L.
; TITLE OF INVENTION: DETECTION OF TRINUCLEOTIDE
; TITLE OF INVENTION: REPEATS
; TITLE OF INVENTION: BY IN SITU HYBRIDIZATION
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0,
; SOFTWARE: Version
; SOFTWARE: #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02861
; FILING DATE: 08 March 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/214,823
; FILING DATE: 17 March 1994
;
; Query Match 5.1%; Score 24; DB 2; Length 31;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5
;
; RESULT 36
; US-08-863-639A-29
; Sequence 29, Application US/08863639A
; Patent No. 5981185
; GENERAL INFORMATION:
; APPLICANT: Matson, Robert S.
; APPLICANT: Coassin, Peter J.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Caskey, C. T.
; TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Corel WordPerfect 8 version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,639A
; FILING DATE: May 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph E. Mueth
; REGISTRATION NUMBER: 20,532
; REFERENCE/DOCKET NUMBER: 11859-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; US-08-863-639A-29
;
; Query Match 5.1%; Score 24; DB 2; Length 33;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Creason, Gary L.
; REGISTRATION NUMBER: 34,310
; REFERENCE/DOCKET NUMBER: 06353/010W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; PCT-US95-02861-14
;
; Query Match 5.1%; Score 24; DB 5; Length 31;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 28 GCTGCTGCTGCTGCTGCTGCTGCT 5
;
; RESULT 36
; US-08-863-639A-29
; Sequence 29, Application US/08863639A
; Patent No. 5981185
; GENERAL INFORMATION:
; APPLICANT: Matson, Robert S.
; APPLICANT: Coassin, Peter J.
; APPLICANT: Rampal, Jang B.
; APPLICANT: Caskey, C. T.
; TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, 9th Floor
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: Corel WordPerfect 8 version
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,639A
; FILING DATE: May 28, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Joseph E. Mueth
; REGISTRATION NUMBER: 20,532
; REFERENCE/DOCKET NUMBER: 11859-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 796-4000
; TELEFAX: (626) 795-6321
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; US-08-863-639A-29
;
; Query Match 5.1%; Score 24; DB 2; Length 33;
; Best Local Similarity 100.0%; Pred. No. 0.014;
; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 3 GCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 37
US-08-068-747-1
; Sequence 1, Application US/08068747
; Patent No. 5695933
; GENERAL INFORMATION:
; APPLICANT: Schalling, Martin
; APPLICANT: Hudson, Thomas J.
; APPLICANT: Housman, David E.
; TITLE OF INVENTION: Direct Determination of Expanded
; TITLE OF INVENTION: Nucleotide Repeats in the Human Genome
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/068,747
; FILING DATE: 28-MAY-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6141
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-068-747-1

Query Match 5.1%; Score 24; DB 1; Length 51;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 3 GCTGCTGCTGCTGCTGCTGCTGCT 26

RESULT 38
US-09-136-080E-50
; Sequence 50, Application US/09136080E
; Patent No. 6518017
; GENERAL INFORMATION:
; APPLICANT: Riley, Timothy A.
; APPLICANT: Brown, Bob D.
; APPLICANT: Arnold, Lyle J.
; TITLE OF INVENTION: COMBINATORIAL ANTISENSE LIBRARY
; FILE REFERENCE: OASBIO.003A
; CURRENT APPLICATION NUMBER: US/09/136,080E
; CURRENT FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50

Query Match 5.1%; Score 24; DB 3; Length 69;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
; LENGTH: 53
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic oligonucleotide
US-09-136-080E-50

Query Match 5.1%; Score 24; DB 4; Length 53;
Best Local Similarity 66.7%; Pred. No. 0.014;
Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 30 GCUGCUGCUGCUGCUGCUGCUGCU 53

RESULT 39
US-09-043-303-14/c
; Sequence 14, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJII, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 57
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-14

Query Match 5.1%; Score 24; DB 3; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 46 GCTGCTGCTGCTGCTGCTGCTGCT 23

RESULT 40
US-09-043-303-13/c
; Sequence 13, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJII, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 69
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-13

Query Match 5.1%; Score 24; DB 3; Length 69;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 40 GCTGCTGCTGCTGCTGCTGCTGCT 17

RESULT 41
US-09-043-303-10/c
; Sequence 10, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-10

Query Match 5.1%; Score 24; DB 3; Length 75;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 25 GCTGCTGCTGCTGCTGCTGCTGCT 2

RESULT 42
US-09-043-303-11/c
; Sequence 11, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-11

Query Match 5.1%; Score 24; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 67 GCTGCTGCTGCTGCTGCTGCTGCT 44

RESULT 43
US-09-043-303-12/c
; Sequence 12, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
```

```
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; EARLIER FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-043-303-12

Query Match 5.1%; Score 24; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 67 GCTGCTGCTGCTGCTGCTGCTGCT 44

RESULT 44
US-09-205-995-75/c
; Sequence 75, Application US/09205995
; Patent No. 6368855
; GENERAL INFORMATION:
; APPLICANT: Xu, Minzhen
; APPLICANT: Qiu, Gang
; APPLICANT: Humphreys, Robert
; TITLE OF INVENTION: CANCER CELL VACCINE
; FILE REFERENCE: U.S. Application 09/205,995, (CIP)
; CURRENT APPLICATION NUMBER: US/09/205,995
; CURRENT FILING DATE: 1998-12-04
; PRIOR APPLICATION NUMBER: 09/036,746
; PRIOR FILING DATE: 1998-03-09
; PRIOR APPLICATION NUMBER: 08/661,627
; PRIOR FILING DATE: 1996-06-11
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 75
; LENGTH: 107
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Reverse gene
; OTHER INFORMATION: construct corresponding to a specific region of
; OTHER INFORMATION: the mouse li gene.
US-09-205-995-75

Query Match 5.1%; Score 24; DB 4; Length 107;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 42 GCTGCTGCTGCTGCTGCTGCTGCT 19

RESULT 45
US-08-469-802B-6/c
; Sequence 6, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ratum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
```

```

; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueiting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueiting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-469-802B-6

Query Match 5.1%; Score 24; DB 1; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCT 77
DB 152 GCTGCTGCTGCTGCTGCTGCT 129

RESULT 45
US-08-267-803B-6/c
; Sequence 6, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harty T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueiting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-267-803B-6

Query Match 5.1%; Score 24; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCT 77
DB 152 GCTGCTGCTGCTGCTGCTGCT 129

RESULT 47
US-08-332-766A-13/c
; Sequence 13, Application US/08332766A
; Patent No. 5843647
; GENERAL INFORMATION:
; APPLICANT: JEFFREYS, Alec J.
; APPLICANT: ARMOUR, John
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,766A
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326052.9
; FILING DATE: 21-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: BIRD, Donald J.
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 160 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-332-766A-13

Query Match 5.1%; Score 24; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 105 GCTGCTGCTGCTGCTGCTGCTGCT 82

RESULT 48
US-09-043-303-17/c
; Sequence 17, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 165
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Probe
US-09-043-303-17

Query Match 5.1%; Score 24; DB 3; Length 165;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 163 GCTGCTGCTGCTGCTGCTGCTGCT 140

RESULT 49
US-08-469-802B-4/c
; Sequence 4, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueeting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
```

```
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-4

Query Match 5.1%; Score 24; DB 1; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 142 GCTGCTGCTGCTGCTGCTGCTGCT 119

RESULT 50
US-08-267-803B-4/c
; Sequence 4, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-267-803B-4

Query Match 5.1%; Score 24; DB 2; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCT 77
Db 142 GCTGCTGCTGCTGCTGCTGCTGCT 119
```

Search completed: May 25, 2004, 19:20:24  
Job time : 94 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2004, 16:30:32 ; Search time 3345 Seconds  
(without alignments)  
6051.180 Million cell updates/sec

Title: SEQ29-NA-TRUNCATED

Perfect score: 467

Sequence: 1 caatgttgctatccacct.....taataaacagttaaaagctg 467

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 10

Total number of hits satisfying chosen parameters: 1846324

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rod.\*

36: em.htg.mam.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	467	100.0	494	6	AX092298	Sequence
2	467	100.0	494	6	AX696953	Sequence
3	467	100.0	495	9	AY358657	Homo sapi
4	467	100.0	2673	9	BC021104	Homo sapi
5	467	100.0	50000	6	AR211704	Sequence
6	467	100.0	151152	9	HS454M7	Human DNA
7	467	100.0	206618	9	AF195953	Homo sapi
8	413	88.4	2243	6	AX877684	Sequence
9	413	88.4	2243	6	BD156791	Primer fo
10	413	88.4	2243	9	AK001855	Homo sapi
11	241	51.6	458	6	AX332625	Sequence
12	241	51.6	458	6	AX332852	Sequence
13	238	51.0	403	6	AX874329	Sequence
14	238	51.0	403	6	BD154391	Primer fo
15	127	27.2	598	9	HUMED38F03	Homo sapi
16	28	6.0	202198	2	AC122549	Mus muscu
17	28	6.0	204171	2	AC131317	Mus muscu
18	27	5.8	227	8	AY023222	Oryza sat
19	27	5.8	373	11	G09506	human STS C
20	27	5.8	1264	8	AB028183	Oryza sat
21	27	5.8	1283	8	AK073848	Oryza sat
22	27	5.8	2000	6	AX654994	Sequence
23	27	5.8	75317	8	SC4357	Sequence
24	27	5.8	88088	8	AP004331	Oryza sat
25	27	5.8	128403	2	AC017472	Oryza sat
26	27	5.8	132059	8	AP003411	Oryza sat
27	27	5.8	143558	2	AC103357	Mus muscu
28	27	5.8	158606	2	AC138056	Medicago
29	27	5.8	163466	3	AC010921	Drosophila
30	27	5.8	170178	5	AL954848	Zebrafish
31	27	5.8	172069	3	AC012160	Drosophila
32	27	5.8	172332	2	AC137828	Medicago
33	27	5.8	177062	10	AC121924	Mus muscu
34	27	5.8	185722	2	AC102553	Mus muscu
35	27	5.8	188834	10	AC098642	Genomic s
36	27	5.8	190402	2	AC114715	Rattus no
37	27	5.8	192539	2	AC132877	Mus muscu
38	27	5.8	192629	2	AC118589	Mus muscu
39	27	5.8	194209	10	AL929249	Mouse DNA
40	27	5.8	195835	10	AL845432	Mouse DNA
41	27	5.8	210174	10	AC124555	Mus muscu
42	27	5.8	211680	10	AL603682	Mouse DNA
43	27	5.8	217278	10	AL604024	Mouse DNA
44	27	5.8	237094	2	AC116661	Mus muscu
45	27	5.8	237332	2	AC131846	Rattus no
46	27	5.8	240501	2	AC099207	Rattus no
47	27	5.8	240561	2	AC129456	Rattus no
48	27	5.8	242978	2	AC093480	Mus muscu
49	27	5.8	250546	2	AC125723	Rattus no
50	27	5.8	258061	2	AC128474	Rattus no
51	27	5.8	263257	2	AC125564	Rattus no
52	27	5.8	264186	2	AC131577	Mus muscu
53	27	5.8	298641	3	AE003504	Drosophila
54	26	5.6	259	10	AF187994	Dipodomys
55	26	5.6	270	3	AY333192	Anopheles
56	26	5.6	339	3	AY333202	Anopheles
57	26	5.6	376	11	G09495	human STS C
58	26	5.6	381	3	AY333201	Anopheles
59	26	5.6	393	3	AY333203	Anopheles
60	26	5.6	393	3	AY333204	Anopheles
61	26	5.6	412	3	AY333205	Anopheles
62	26	5.6	423	3	AY333206	Anopheles
63	26	5.6	457	8	FVE508245	Fragaria
64	26	5.6	502	8	AY158013	Fucaleae
65	26	5.6	735	10	BC048870	Mus muscu

Pred. No. is the number of results predicted by chance to have a

```

66 5.6 1022 10 MMU080893 U08093 Mus musculus
67 5.6 1048 10 BC012681 BC012681 Mus muscu
68 5.6 1156 10 MMU08089 U08089 Mus muscu
69 5.6 1232 3 AY118433 AY118433 Drosophi
70 5.6 1685 10 BC036332 BC036332 Mus muscu
71 5.6 3063 3 AB060995 AB060995 Drosophi
72 5.6 7072 3 AB090813 AB090813 Anophele
73 5.6 7183 2 AC012983 AC012983 Drosophi
74 5.6 7673 10 AF026032 AF026032 Mus muscu
75 5.6 14532 2 AC019984 AC019984 Drosophi
76 5.6 15911 3 AF096709 AF096709 Drosophi
77 5.6 25724 2 AC019732 AC019732 Drosophi
78 5.6 36328 3 LMFL5213 LMFL5213 Leishmani
79 5.6 37891 2 AC014982 AC014982 Drosophi
80 5.6 38418 3 AY190934 AY190934 Drosophi
81 5.6 45300 8 AP005115 AP005115 Oryza sat
82 5.6 45375 5 AF328738 AF328738 Agelaius
83 5.6 46427 3 AY190956 AY190956 Drosophi
84 5.6 46904 2 AC100911 AC100911 Mus muscu
85 5.6 56366 2 AC019800 AC019800 Drosophi
86 5.6 65464 2 AC121129 AC121129 Mus muscu
87 5.6 78456 10 AL627393 AL627393 Mouse DNA
88 5.6 78964 2 AC020460 AC020460 Drosophi
89 5.6 79786 6 AX695800 AX695800 Sequence
90 5.6 95946 2 AC138274 AC138274 Mus muscu
91 5.6 100500 2 AC133924 AC133924 Magnapor
92 5.6 110000 2 AC110168_1 Continuation (2 of
93 5.6 110000 2 AC110168_2 Continuation (3 of
94 5.6 110000 2 AC110642_2 Continuation (3 of
95 5.6 110000 2 AC110959_0 AC110959 Rattus no
96 5.6 110000 2 AC116234_1 Continuation (2 of
97 5.6 110000 2 LMFLCHR31_18 Continuation (19 o
98 5.6 110325 8 AP005003 AP005003 Oryza sat
99 5.6 113887 10 BX539328 BX539328 Mouse DNA
100 5.6 114135 2 AC137652 AC137652 Bos tauru

```

## ALIGNMENTS

```

RESULT 1
AX092298 494 bp DNA linear PAT 21-MAR-2001
LOCUS Sequence 29 from Patent WO0116318.
DEFINITION AX092298
ACCESSION AX092298
VERSION AX092298.1 GI:13444463
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ferrera, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,
Eaton, D.L., Gao, W.Q., Fan, J., Hotstein, D., Fong, S., Goddard, A.,
Godowski, P.J., Grimaldi, C.J., Gurney, A.L., Tamas, D., Wood, W.I.,
Grimaldi, C.J., Hillan, K.J., Paoni, N.F., Roy, M.A. and Watanabe, C.K.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0116318-A 29 08-MAR-2001;
Genentech, Inc. (US)
FEATURES
source
location/Qualifiers
1..494
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATGTTTGCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
Db |
1 CAATGTTTGCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
QY 1 CAATGTTTGCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 180
Db |
1 CAATGTTTGCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 180

```

```

QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGAGCTGCTGCGTGCCAGAAA 120
Db |
61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGAGCTGCTGCGTGCCAGAAA 120
QY 121 GTCTCTTCTGCCACTCAACGCCCTCCATCAGGATTTGGGCTTTCTTCCCTCCCTTCTG 180
Db |
121 GTCTCTTCTGCCACTCAACGCCCTCCATCAGGATTTGGGCTTTCTTCCCTCCCTTCTG 180
QY 181 TGTCTCTGCTCATCGGCTGCGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 240
Db |
181 TGTCTCTGCTCATCGGCTGCGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 240
QY 241 GAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAAAGAGGATAGTGGGTG 300
Db |
241 GAGAAAGTGGGGATGGCTTAAGAAAGCTGGAGATAGGGAACAGAAAGAGGATAGTGGGTG 300
QY 301 GGTCTAGGGGGCTGCTCTTAAAGTGGTGTGTTATGATTTCTTATCTTAACTATACAA 360
Db |
301 GGTCTAGGGGGCTGCTCTTAAAGTGGTGTGTTATGATTTCTTATCTTAACTATACAA 360
QY 361 AGATATTAAGGCGCTGCTTCAATTAAGAAATGTTCCCTCCCTGCTGCTCAATGTTGTAA 420
Db |
361 AGATATTAAGGCGCTGCTTCAATTAAGAAATGTTCCCTCCCTGCTGCTCAATGTTGTAA 420
QY 421 AGATGTTTCTGTTAAATGATGCTTTTATAATAACAGTTTAAAGCTG 467
Db |
421 AGATGTTTCTGTTAAATGATGCTTTTATAATAACAGTTTAAAGCTG 467
RESULT 2
AX696953 494 bp DNA linear PAT 02-APR-2003
LOCUS Sequence 21 from Patent WO0078961.
DEFINITION AX696953
ACCESSION AX696953
VERSION AX696953.1 GI:29497962
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Ferrera, N., Stewart, T.A., Williams, P.M., Baker, K.P., Desnoyers, L.,
Eaton, D.L., Gao, W.Q., Fan, J., Hotstein, D., Fong, S., Goddard, A.,
Godowski, P.J., Gurney, A.L., Smith, V., Tamas, D., Wood, W.I.,
Grimaldi, C.J., Hillan, K.J., Paoni, N.F., Roy, M.A. and Watanabe, C.K.
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding
the same
JOURNAL Patent: WO 0078961-A 21 28-DEC-2000;
Genentech, Inc. (US)
FEATURES
source
location/Qualifiers
1..494
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 467; DB 6; Length 494;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATGTTTGCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
Db |
1 CAATGTTTGCCCTATCACCTCCCAAGCCCTTTACCTATGCTGCTGCTAACGCTGCTG 60
QY 61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGAGCTGCTGCGTGCCAGAAA 120
Db |
61 CTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGAGTGGGAGCTGCTGCGTGCCAGAAA 120
QY 121 GTCTCTTCTGCCACTCAACGCCCTCCATCAGGATTTGGGCTTTCTTCCCTCCCTTCTG 180
Db |
121 GTCTCTTCTGCCACTCAACGCCCTCCATCAGGATTTGGGCTTTCTTCCCTCCCTTCTG 180

```

```

QY 181 TGTCTCTGCTCATCGGCGCTGCATGACCTGCAGCCCAAGCCAGCCCGCGGGAGGG 240
Db 181 TGTCTCTGCTCATCGGCGCTGCATGACCTGCAGCCCAAGCCCGCGGGAGGG 240
QY 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAAAGAGGTAGTGGGTG 300
QY 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 360
Db 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 360
QY 361 AGATATTAAAGCCCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 420
Db 361 AGATATTAAAGCCCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 420
QY 421 AGATTGTTCTGTGTAATATGCTTTTAAATTAACAGTTAAAGCTG 467
Db 421 AGATTGTTCTGTGTAATATGCTTTTAAATTAACAGTTAAAGCTG 467

RESULT 3
AY358657 495 bp mRNA linear PRI 03-OCT-2003
LOCUS Homo sapiens clone DNA56862 APELIN (UNQ471) mRNA, complete cds.
DEFINITION
ACCESSION AY358657
VERSION AY358657.1 GI:37182435
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J.,
Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B.,
Dowd,J., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E.,
Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jih,Y., Johnson,S.,
Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C.,
Schoenfeld,J., Seshagiri,S., Simmons,L., Singh,J., Smith,V.,
Stinson,J., Vagts,A., Vanden,R., Watanabe,C., Wleand,D., Woods,K.,
Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z.,
Goddard,A., Wood,W.I. and Godowski,P.
The Secreted Protein Discovery Initiative (SPDI), a Large-Scale
Effort to Identify Novel Human Secreted and Transmembrane Proteins:
A Bioinformatics Assessment
Genome Res. 13 (10), 2265-2270 (2003)
12975309
2 (bases 1 to 495)
Clark,H.F.
Direct Submission
Submitted (01-AUG-2003) Department of Bioinformatics, Genentech,
Inc., 1 DNA Way, South San Francisco, CA 94080, USA
FEATURES
source
1. .495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="DNA56862"
1. .495
/locus_tag="UNQ471"
40. .261
/locus_tag="UNQ471"
/notes="PRO831"
/codon_start=1
/product="APELIN"
/protein_id="AAQ89020.1"
/db_xref="GI:37182436"
/translations="MLLLTLLLLLLKSGEWLGVAGKVSSATDAPIDWAFPP
SFLLCLLPHRPAMTCSQAQPRGEKVGDS"

ORIGIN
Query Match 100.0%; Score 467; DB 9; Length 495;
Best Local Similarity 100.0%; Pred. No. 2.3e-258;


```

```

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTTGGCTATPCACCTCCCAAGCCCTTTACCTATGCTGCTCTAACGCTGCTG 60
Db 1 CAATGTTTGGCTATPCACCTCCCAAGCCCTTTACCTATGCTGCTCTAACGCTGCTG 60
QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTTGAAGTGGGACTGCTGCTGCCAGAAA 120
Db 61 CTGCTGCTGCTGCTTAAAGCTCATGCTTGAAGTGGGACTGCTGCTGCCAGAAA 120
QY 121 GTCTCTTCTCCCACTGACGCGCCCATCAGGATTTGGGCTTCTTTCCCTCCCTTCTTCTG 180
Db 121 GTCTCTTCTCCCACTGACGCGCCCATCAGGATTTGGGCTTCTTTCCCTCCCTTCTTCTG 180
QY 181 TGCTCTGCTGCTATCGGCTGCTGATGACCTGAGCCCAAGCCCGCTGGGGAGGG 240
Db 181 TGCTCTGCTGCTATCGGCTGCTGATGACCTGAGCCCAAGCCCGCTGGGGAGGG 240
QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGGGGTAGTGGGTG 300
Db 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAGGGGTAGTGGGTG 300
QY 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 360
Db 301 GGCTAGGGGGGCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 360
QY 361 AGATATTAAAGCCCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 420
Db 361 AGATATTAAAGCCCTGCTTATTTAAAGTGGTGTATGATCTTATGATCTTATCTAATTTATACAA 420
QY 421 AGATTGTTCTGTGTAATATGCTTTTAAATTAACAGTTAAAGCTG 467
Db 421 AGATTGTTCTGTGTAATATGCTTTTAAATTAACAGTTAAAGCTG 467

RESULT 4
BC021104
LOCUS Homo sapiens apelin, mRNA (cDNA clone MGC:31846 IMAGE:4586949),
complete cds.
DEFINITION BC021104
ACCESSION BC021104.1 GI:18088893
VERSION BC021104
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22386257
12477932
REFERENCE 2 (bases 1 to 2673)
AUTHORS Strausberg,R.

```

**TITLE**  
JOURNAL

**REMARK**  
COMMENT

Direct Submission  
Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue: Procurement: DCID/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garlang, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 40 Row: f Column: 18  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 21314667.

**FEATURES**  
source  
Location/Qualifiers  
1. .2673  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:31846 IMAGE:4586949"  
/tissue\_type="Kidney, renal cell adenocarcinoma"  
/clone\_lib="NIH MGC 14"  
/lab\_host="DH10B-R"  
/notes="Vector: pOTB7"  
1. .2673  
/gene="APLN"  
/db\_xref="LocusID:8862"  
420. .788  
/db\_xref="MIN:300297"  
/codon\_start=1  
/product="apelin"  
/protein\_id="AAH21104.1"  
/db\_xref="GI:18088594"  
/db\_xref="LocusID:8862"  
/translation="MSATWSPRQGMQGGPGREVGNSSAASGPASPIRDPCLSEAGL KGPSPAPRRLLGLHLRVCFSGGLTSIQLSPRTCCSHQWQLFSPACFPQWRAPGCSL DDSRLTRIRPHLPSPSLD"  
ORIGIN  
Query Match 100.0%; Score 467; DB 9; Length 2673;  
Best Local Similarity 100.0%; Pred. No. 2.4e-258;  
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAATGTTTGGCTATCCACCTCCCGAGCCCTTTACCTATGCTGCTGCTAAGCTGCTG 60  
Db 2188 CAATGTTTGGCTATCCACCTCCCGAGCCCTTTACCTATGCTGCTGCTAAGCTGCTG 2247  
QY 61 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Db 2248 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2307  
QY 121 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
Db 2308 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2367  
QY 181 TGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Db 2368 TGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2427

QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTACTGGGTG 300  
Db 2428 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATAGGGAACAGAGAGGGTACTGGGTG 2487  
QY 301 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATTCTTATATCTAATTTATCAA 360  
Db 2488 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATTCTTATATCTAATTTATCAA 2547  
QY 361 AGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGTTGTTCAATGTTGTAA 420  
Db 2548 AGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGTTGTTCAATGTTGTAA 2607  
QY 421 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTAAAGCTG 467  
Db 2608 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTAAAGCTG 2654

RESULT 5  
AR211704/c  
LOCUS AR211704  
DEFINITION Sequence 4 from patent US 6399349.  
ACCESSION AR211704  
VERSION AR211704.1 GI:21515097  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 50000)  
AUTHORS Ryan, J.W. and Sprinkle, T.Joe.Curtis.  
TITLE Human aminopeptidase p gene  
JOURNAL Patent: US 6399349-A 4 04-JUN-2002;  
FEATURES Location/Qualifiers  
source  
1. .50000  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Query Match 100.0%; Score 467; DB 6; Length 50000;  
Best Local Similarity 100.0%; Pred. No. 2.6e-258;  
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAATGTTTGGCTATCCACCTCCCGAGCCCTTTACCTATGCTGCTGCTAAGCTGCTG 60  
Db 1518 CAATGTTTGGCTATCCACCTCCCGAGCCCTTTACCTATGCTGCTGCTAAGCTGCTG 1459  
QY 61 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Db 1458 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1399  
QY 121 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180  
Db 1398 GTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1339  
QY 181 TGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
Db 1338 TGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1279  
QY 241 GAGAAAGTGGGGATGGCTTAAAGAGCTGGGAGATAGGGAACAGAGAGGGTATGTTGTTG 300  
Db 1278 GAGAAAGTGGGGATGGCTTAAAGAGCTGGGAGATAGGGAACAGAGAGGGTATGTTGTTG 1219  
QY 301 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATTCTTATATCTAATTTATCAA 360  
Db 1218 GGCTAGGGGGCTGCTTATTTAAAGTGGTGTGTTTATGATTCTTATATCTAATTTATCAA 1159  
QY 361 AGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGTTGTTCAATGTTGTAA 420  
Db 1158 AGATATTAAAGGCCCTGTTTCATTAAAGAAATGTTCCCTTCCCTGTTGTTCAATGTTGTAA 1099  
QY 421 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTAAAGCTG 467  
Db 1098 AGATTGTTCTGTGTAATAATGCTTTTATAATAACAGTTAAAGCTG 1052



```
repeat_region /note="MER33 repeat: matches 1. .323 of consensus"
7178. 8315
repeat_region /note="L1P22 repeat: matches 5000. .6146 of consensus"
11245. .11294
repeat_region /note="MER5B repeat: matches 109. .178 of consensus"
11345. .11657
repeat_region /note="AluYb8 repeat: matches 1. .310 of consensus"
11736. .11797
repeat_region /note="31 copies 2 mer tt 72% conserved"
12436. 12562
repeat_region /note="L2 repeat: matches 2579. .2710 of consensus"
13276. .13457
repeat_region /note="AluSg/x repeat: matches 129. .310 of consensus"
13679. .13873
repeat_region /note="MIR repeat: matches 5. .213 of consensus"
14132. 14206
repeat_region /note="MIR repeat: matches 163. .233 of consensus"
14207. .14413
repeat_region /note="MER8 repeat: matches 2. .239 of consensus"
14414. .14548
repeat_region /note="MIR repeat: matches 12. .163 of consensus"
14793. .15073
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
15383. .15589
repeat_region /note="MER3 repeat: matches 1. .209 of consensus"
16139. .16279
repeat_region /note="MER5B repeat: matches 29. .178 of consensus"
16317. .16466
repeat_region /note="MIR repeat: matches 52. .212 of consensus"
16667. .17000
repeat_region /note="L1P49 repeat: matches 5829. .6163 of consensus"
17001. .17301
repeat_region /note="AluYb8 repeat: matches 1. .302 of consensus"
17302. .17642
repeat_region /note="L1P29 repeat: matches 5491. .5829 of consensus"
18285. .18620
repeat_region /note="2 copies 168 mer 78% conserved"
18772. .18981
repeat_region /note="MIR repeat: matches 6. .242 of consensus"
18945. .19005
repeat_region /note="L2 repeat: matches 2648. .2702 of consensus"
20349. 20489
repeat_region /note="MIR repeat: matches 1. .150 of consensus"
20543. .20697
repeat_region /note="MER5A repeat: matches 22. .188 of consensus"
21496. .21830
repeat_region /note="L1MC4 repeat: matches 7477. .7849 of consensus"
22420. 22650
repeat_region /note="MER46A repeat: matches 1. .236 of consensus"
22894. .22965
repeat_region /note="3 copies 24 mer 83% conserved"
24309. .24606
repeat_region /note="AluX repeat: matches 1. .299 of consensus"
25081. .25134
repeat_region /note="27 copies 2 mer ta 70% conserved"
25168. .25193
repeat_region /note="13 copies 2 mer tg 100% conserved"
25168. .25193

Query Match 100.0%; Score 467; DB 9; Length 151152;
Best Local Similarity 100.0%; Pred. No. 2,7e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATGTTTGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTGCTAACGCTGCTG 60
Db 89718 CAATGTTTGCTATCCACCTCCCAAGCCCTTACCTATGCTGCTGCTAACGCTGCTG 89659

QY 61 CTGCTGCTGCTGCTTAAAGCTCATGCTTGAGTGGGACTGCTGCTGCCAGAAA 120
Db 89658 CTGCTGCTGCTGCTTAAAGCTCATGCTTGAGTGGGACTGCTGCTGCCAGAAA 89599

QY 121 GTCTCTTCTGCCACTGACGCCCCCATCAGGATGGGCCCTTCTTCCCTTCTCTTCTG 180
Db 89598 GTCTCTTCTGCCACTGACGCCCCCATCAGGATGGGCCCTTCTTCTTCTCTTCTG 89539
```

```
QY 181 TGTCTCTGCTCATCGGCTGCGACCTGCGACCGCAAGCCCGGCGGAAGGG 240
Db 89538 TGTCTCTGCTCATCGGCTGCGACCTGCGACCGCAAGCCCGGCGGAAGGG 89479

QY 241 GAGAAAGTGGGGATGCTAAGAAAGCTGGAGATAGGGAACAGAGGGTAGTGGGTG 300
Db 89478 GAGAAAGTGGGGATGCTAAGAAAGCTGGAGATAGGGAACAGAGGGTAGTGGGTG 89419

QY 301 GGCTAGGGGGCTGCCCTTATTAAAGTGGTGTGTTATGATCTTATCTAATTTATACAA 360
Db 89418 GGCTAGGGGGCTGCCCTTATTAAAGTGGTGTGTTATGATCTTATCTAATTTATACAA 89359

QY 361 AGATATTAAAGCCCTGCTCATTAAGAAATGTTCCCTCCCTGTTCAATGTTTGTAA 420
Db 89358 AGATATTAAAGCCCTGCTCATTAAGAAATGTTCCCTCCCTGTTCAATGTTTGTAA 89299

QY 421 AGATTGTTCTGTTAAATATGCTTTATATAAACAAGTTAAAAAGCTG 467
Db 89298 AGATTGTTCTGTTAAATATGCTTTATATAAACAAGTTAAAAAGCTG 89252

RESULT 7
AF195953/c 206618 bp DNA linear PRI 26-MAR-2002
LOCUS Homo sapiens membrane-bound aminopeptidase P (XNPEP2) gene,
DEFINITION complete cds.
ACCESSION AF195953
VERSION AF195953.2 GI:19718557
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 206618)
AUTHORS Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C.
TITLE Human membrane-bound aminopeptidase P genomic DNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 206618)
AUTHORS Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Vascular Biology Center, Medical College of
Georgia, 1120 15th Street, Augusta, GA 30912, USA
REFERENCE 3 (bases 1 to 206618)
AUTHORS Ryan,J.W., Jin,L., Horvath,I. and Sprinkle,T.J.C.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2002) Vascular Biology Center, Medical College of
Georgia, 1120 15th Street, Augusta, GA 30912, USA
REMARK Sequence update by submitter
COMMENT On Mar 26, 2002 this sequence version replaced gi:11066156.
FEATURES
Location/Qualifiers
1..206618
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
144189..176791
/gene="XNPEP2"
join(144189..144502,147348..147421,149203..149313,
150445..150508,151478..151582,151837..151923,
152849..152995,155710..155811,156987..157068,
157392..157587,158401..158490,159714..159823,
160536..160613,161726..161797,164422..164482,
165754..165823,166414..166518,167248..167307,
167936..168012,172845..172934,173533..176791)
/gene="XNPEP2"
/product="membrane-bound aminopeptidase P"
144189..144453
/gene="XNPEP2"
join(144454..144502,147348..147421,149203..149313,
150445..150508,151478..151582,151837..151923,
152849..152995,155710..155811,156987..157068,
157392..157587,158401..158490,159714..159823,
160536..160613,161726..161797,164422..164482,
165754..165823,166414..166518,167248..167307,
167936..168012,172845..172934,173533..176791)
/gene="XNPEP2"
5'UTR
144189..144453
/gene="XNPEP2"
CDS
join(144454..144502,147348..147421,149203..149313,
150445..150508,151478..151582,151837..151923,
152849..152995,155710..155811,156987..157068,
157392..157587,158401..158490,159714..159823,
160536..160613,161726..161797,164422..164482,
165754..165823,166414..166518,167248..167307,
167936..168012,172845..172934,173533..176791)
/gene="XNPEP2"
```

```
165754. .165823,166414. .166518,167248. .167307,
167936. .168012,172845. .172934,173533. .173727)
/gene="XNPEP2"
/codon_start=1
/product="membrane-bound aminopeptidase p"
/db_xref="GI:11066157"
/translation="MARAHWGCCPWLVLCCAMGHKTPVDLGGQVNCSTNPPYLP
VTVNTMSLTALQOMCTONLSAVIPGTAHMEYVIGQDERRAMITGRTSGAGIA
VVTWKAAWDSYRWYQAEQMCNWEHKEVGTPIVTLWLTETPAGGVRGDFDPL
LSDTWESYDALQSNQQLVITNLVDLVWGERPPVNPQPIYALQEAFTGSTWQE
KSVGRSQMQHQVPTAVLLSALAEETAWLFLRASDIPYFPFFSYTLTDSIRLF
ANKRSFSETSYLNSCTGPMVCQIEDYSQVRDSIQAYSIGDVRIWLTGTYMYGIV
EMIPEKLVITDTPYSPVMTKAVKNSKEQALLKASHVRDAVAVIRVWLEKNVPKTV
DEFGAETVDFRGEQSPSPETISASGLNALAHYSPKELNKLSDENYLLD
SGGYWDTGTTDITVHWGTSPAFQKAYTVLIGNDLSLIFPAITSGRMWAFAR
RALNDAGLNYGHGTGHGIGNFLCVHEPVGFSNNIAWAKGMFTISPEFYKDGEGFI
RLEDVALVVEAKTIPGSLYTFEVVSPVYDRNLIDVLSLSPHELOLYNLRYQTIREK
VGPBLQRQLLEEFEMIQOHTPELAARAPDTASWASVLVWSTLAILGMSV"
/misc_feature
160606. .160608
/gene="XNPEP2"
/notes="Putative proton shuttle; unclassified site"
/misc_feature
161778. .161780
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
164435. .164437
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
166473. .166475
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
167936. .167937
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
167977. .167979
/gene="XNPEP2"
/notes="divalent metal ligand; metal-binding site"
3'UTR
175728. .176791
/gene="XNPEP2"
ORIGIN
Query Match 100.0%; Score 467; DB 9; Length 206618;
Best Local Similarity 100.0%; Pred. No. 2.8e-258;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAATGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTAAAGCGTCTG 60
DB 51518 CAATGTTTGCTATCCACCTCCCCCAAGCCCTTTACCTATGCTGCTAAAGCGTCTG 51459
QY 61 CTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGGAAGTGGTGGTGGCCAGAA 120
DB 51458 CTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGGAAGTGGTGGTGGCCAGAA 51399
QY 121 GTCTCTTCTGCCACTGACGCCCCATCAGGAGTTGGGCCTCTTTTCCCTCTCTTTCTG 180
DB 51398 GTCTCTTCTGCCACTGACGCCCCATCAGGAGTTGGGCCTCTTTTCCCTCTCTTTCTG 51339
QY 181 TGTCTCTGCTCATCGGCTGCGCATGCTGACGCCCAAGCCAGCCCGTGGGGAAGG 240
DB 51338 TGTCTCTGCTCATCGGCTGCGCATGCTGACGCCCAAGCCAGCCCGTGGGGAAGG 51279
QY 241 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATGGGGAAGAGAGAGGTAGTGGGTG 300
DB 51278 GAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATGGGGAAGAGAGAGGTAGTGGGTG 51219
QY 301 GGCTAGGGGGCTGCCTATTATTAAGTGGTGGTATGATTCCTTACTAATTTATACAA 360
DB 51218 GGCTAGGGGGCTGCCTATTATTAAGTGGTGGTATGATTCCTTACTAATTTATACAA 51159
QY 361 AGATATTAAAGGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 420
DB 51158 AGATATTAAAGGCCCTGTTCAATTAAGAAATGTTCCCTTCCCTGCTGTTCAATGTTGTAA 51099
QY 421 AGATTGTTCTGTAAATATATCTTTATAATAAACAAGTTAAAGCTG 467
DB 51098 AGATTGTTCTGTAAATATATCTTTATAATAAACAAGTTAAAGCTG 51052
RESULT 8
AX877684
LOCUS AX877684 2243 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 12589 from Patent EP1074617.
ACCESSION AX877684
VERSION AX877684.1 GI:40032420
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 12589 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
Location/Qualifiers
1..2243
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
6..374
/notes="unnamed protein product"
/contig_start=1
/protein_id="CA89834.1"
/db_xref="GI:40032421"
/translation="NSRATMCSPEGQMGQPGREVGNSSAASPASIRNPCLSEAGL
XGPPSAHPLRLCLLHLVCFSGGLTSLQSPRTCCSHQWQLFSPACFPQWRAPGCSL
DSSRLTRINPVLPGPSLD"
ORIGIN
Query Match 88.4%; Score 413; DB 6; Length 2243;
Best Local Similarity 100.0%; Pred. No. 4.2e-227;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 55 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCC 114
DB 1831 CTGCTGCTGCTGCTGCTGCTTAAAGGCTCATGCTTGGAGTGGGACTGGTGGTGGCC 1890
QY 115 CAGAAAGTCTTCTGCCACTGACGCCCCATCAGGAGTTGGGCCTTTTTCCTCCCTTC 174
DB 1891 CAGAAAGTCTTCTGCCACTGACGCCCCATCAGGAGTTGGGCCTTTTTCCTCCCTTC 1950
QY 175 TTTTCTGCTGCTGCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 234
DB 1951 TTTTCTGCTGCTGCTGCTCATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2010
QY 235 GAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATGGGGAAGAGAGGTAG 294
DB 2011 GAAGGGGAGAAAGTGGGGATGGCTAAGAAAGCTGGGAGATGGGGAAGAGAGGTAG 2070
QY 295 TGGGTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGG 354
DB 2071 TGGGTGGGCTAGGGGGCTGCTTATTTAAAGTGGTGGTGGTGGTGGTGGTGGTGG 2130
QY 355 ATACAAGATATTAGGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTG 414
DB 2131 ATACAAGATATTAGGGCCCTGCTTCAATTAAGAAATGTTCCCTTCCCTGCTGCTG 2190
QY 415 TTGTAAGATTTGTTCTGTTAAATATGCTTTTATAAACAAGTTAAAAAGCTG 467
DB 2191 TTGTAAGATTTGTTCTGTTAAATATGCTTTTATAAACAAGTTAAAAAGCTG 2243
RESULT 9
BD156791
LOCUS BD156791 2243 bp DNA linear PAT 17-JAN-2003
```





```

Db      2131 ATACAAAGATATTAAGGCCCTGTTTCATTAAGAAATGTTCCCTCCCTGTTGTTCAATGT 2190
QY      415 TTGTAAGATTTGTTCTGTGTAATAATGCTTTATATAACAGTTAAAGCTG 467
Db      2191 TTGTAAGATTTGTTCTGTGTAATAATGCTTTATATAACAGTTAAAGCTG 2243

RESULT 11
LOCUS   AX332625/2 458 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3134 from Patent WO0194629.
ACCESSION AX332625
VERSION   AX332625.1 GI:18123259
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
          Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE    Cancer gene determination and therapeutic screening using signature
          gene sets
JOURNAL  Avalon Pharmaceuticals (US)
FEATURES
source   1..458
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.6%; Score 241; DB 6; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.4e-127;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CCCGTGGGGAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 286
Db 252 CCCGTGGGGAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 193
QY 287 GAGGTTAGTGGGTGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 346
Db 192 GAGGTTAGTGGGTGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 133
QY 347 ACTAATTTATACAAAGATATTAAGGCCCTGTTCAATTAAGAAATGTTCCCTCCCTGTG 406
Db 132 ACTAATTTATACAAAGATATTAAGGCCCTGTTCAATTAAGAAATGTTCCCTCCCTGTG 73
QY 407 TTCAATGTTTGTAAAGATTTCTGTGTAAATATGCTTTATATAAACAAGTTAAAGCT 466
Db 72 TTCAATGTTTGTAAAGATTTCTGTGTAAATATGCTTTATATAAACAAGTTAAAGCT 13
QY 467 G 467
Db 12 G 12

RESULT 12
LOCUS   AX332852/2 458 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3361 from Patent WO0194629.
ACCESSION AX332852
VERSION   AX332852.1 GI:18123486
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
          Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE    Cancer gene determination and therapeutic screening using signature
          gene sets
JOURNAL  Avalon Pharmaceuticals (US)
FEATURES
source   1..458
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.0%; Score 238; DB 6; Length 403;
Best Local Similarity 99.7%; Pred. No. 7.5e-126;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 TGTGTCTCTCTCATCGGCTGCGCATGACCTGCAGCAAGCCAGCCCGCTGGGGAAG 238
Db 289 TGTGTCTCTCTCATCGGCTGCGCATGACCTGCAGCAAGCCAGCCCGCTGGGGAAG 230
QY 239 GGGAGAAAGTGGGGATGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTTAGTGG 298
Db 229 GGGAGAAAGTGGGGATGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTTAGTGG 170
QY 299 TGGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATCTTATATAAATTATAC 358
Db 169 TGGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATCTTATATAAATTATAC 110

```

```

JOURNAL Patent: WO 0194629-A 3361 13-DEC-2001;
          Avalon Pharmaceuticals (US)
FEATURES
source   1..458
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.6%; Score 241; DB 6; Length 458;
Best Local Similarity 100.0%; Pred. No. 1.4e-127;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 CCCGTGGGGAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 286
Db 252 CCCGTGGGGAAGGGGAGAAAGTGGGGGATGGCTAAGAAAGCTGGAGATAGGGAACAGAA 193
QY 287 GAGGTTAGTGGGTGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 346
Db 192 GAGGTTAGTGGGTGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATTCCTAT 133
QY 347 ACTAATTTATACAAAGATATTAAGGCCCTGTTCAATTAAGAAATGTTCCCTCCCTGTG 406
Db 132 ACTAATTTATACAAAGATATTAAGGCCCTGTTCAATTAAGAAATGTTCCCTCCCTGTG 73
QY 407 TTCAATGTTTGTAAAGATTTCTGTGTAAATATGCTTTATATAAACAAGTTAAAGCT 466
Db 72 TTCAATGTTTGTAAAGATTTCTGTGTAAATATGCTTTATATAAACAAGTTAAAGCT 13
QY 467 G 467
Db 12 G 12

RESULT 13
LOCUS   AX874329/c 403 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 9234 from Patent EP1074617.
ACCESSION AX874329
VERSION   AX874329.1 GI:40029100
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
          Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE    Primers for synthesising full-length cDNA and their use
JOURNAL  Patent: EP 1074617-A 9234 07-FEB-2001;
          Research Association for Biotechnology (JP)
FEATURES
source   1..403
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Query Match 51.0%; Score 238; DB 6; Length 403;
Best Local Similarity 99.7%; Pred. No. 7.5e-126;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 179 TGTGTCTCTCTCATCGGCTGCGCATGACCTGCAGCAAGCCAGCCCGCTGGGGAAG 238
Db 289 TGTGTCTCTCTCATCGGCTGCGCATGACCTGCAGCAAGCCAGCCCGCTGGGGAAG 230
QY 239 GGGAGAAAGTGGGGATGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTTAGTGG 298
Db 229 GGGAGAAAGTGGGGATGCTTAAGAAAGCTGGAGATAGGGAACAGAGAGGTTAGTGG 170
QY 299 TGGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATCTTATATAAATTATAC 358
Db 169 TGGGCTAGGGGGCTGCCCTTATTTAAAGTGGTGTGTTATGATCTTATATAAATTATAC 110

```

```

QY 359 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 418
Db 109 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 50
QY 419 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 467
Db 49 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 50

RESULT 14
BD154391/c
LOCUS BD154391 403 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD154391
VERSION BD154391.1 GI:27860149
KEYWORDS JP 2002191363-A/9234.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 403)
AUTHORS Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 9234 09-JUL-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002191363-A/9234
PD 09-JUL-2002
PF 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH key
Location/Qualifiers
FT source 1..403
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..403
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 51.0%; Score 238; DB 6; Length 403;
Best Local Similarity 99.7%; Pred. No. 7.5e-126;
Matches 288; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 179 TGTGTCCTCTGCTCATGCGCCGCGCATGACCTGCGAGCCAGCCCGCTGGGGAAG 238
Db 289 TGTGTCCTCTGCTCATGCGCCGCGCATGACCTGCGAGCCAGCCCGCTGGGGAAG 230
QY 239 GGGAGAAAGTGGGGGATGGCTAGAAAGCTGGGAGATGGGACAGAGGGTAGTGGG 298
Db 229 GGGAGAAAGTGGGGGATGGCTAGAAAGCTGGGAGATGGGACAGAGGGTAGTGGG 170
QY 299 TGGGCTAGGGGGCTCCCTTAATTAAGTGGTGTATGATCTTATCAATTAATTATAC 358
Db 169 TGGGCTAGGGGGCTCCCTTAATTAAGTGGTGTATGATCTTATCAATTAATTATAC 110
QY 359 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 418
Db 109 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 50
QY 419 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 467
Db 49 AAAGATTATTAAGGCGCTCTTCATTAAAGAAATGTTCCCTCCCTGCTGTTCAATGTTGT 50

```

```

RESULT 15
HUMZD38F03
LOCUS HUMZD38F03 598 bp mRNA linear PRI 29-AUG-1998
DEFINITION Homo sapiens full length insert cDNA clone ZD38F03.
ACCESSION AF086248
VERSION AF086248.1 GI:3483593
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 598)
AUTHORS Moessner,J., Tan,F., Marra,M., Kucaba,T., Yandell,M., Martin,J.,
Marth,G., Bowles,L., Wylie,T., Bowers,Y., Steptoe,M., Theising,B.,
Geisel,S., Allen,M., Underwood,K., Chappell,J., Person,B.,
Gibbons,M., Harvey,N., Pape,D., Chamberlain,A., Morales,R.,
Schurk,R., Ritter,E., Kohn,S., Swaller,T., Behymer,K., Hillier,L.,
Wilson,R. and Waterston,R.
TITLE Full Clone Sequencing of the Longest Available Member from Each
Unigene Cluster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 598)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:est@watson.wustl.edu

```

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

```

FEATURES
source
Location/Qualifiers
1..598
/organism='Homo sapiens'
/mol_type='mRNA'
/db_xref='taxon:9606'
/clone='IMAGE:342941'
/clone_lib='Soares_fetal_heart_NbHH19W'

```

ORIGIN

```

Query Match 27.2%; Score 127; DB 9; Length 598;
Best Local Similarity 100.0%; Pred. No. 1.3e-61;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 341 TCTTATCTAATTTATACAAAGATTAAGCCCTGTTTCATTAGAAATTTGTCCTTCC 400
Db 1 TCTTATCTAATTTATACAAAGATTAAGCCCTGTTTCATTAGAAATTTGTCCTTCC 60
QY 401 CCGTGTTTCAATGTTTGTAAAGATTTGTTCTGTGTAATATGTTCTTTATAATAACAGTTA 460
Db 61 CCGTGTTTCAATGTTTGTAAAGATTTGTTCTGTGTAATATGTTCTTTATAATAACAGTTA 120
QY 461 AAAGCTG 467
Db 121 AAAGCTG 127

```

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information  
 Center project name: L23913  
 Center clone name: 406\_O\_24

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 199596 bases at least Q40  
 Consensus quality: 200632 bases at least Q30  
 Consensus quality: 201122 bases at least Q20  
 Insert size: 183000; agarose-fp  
 Insert size: 201398; sum-of-contigs  
 Quality coverage: 10.7 in Q20 bases; agarose-fp  
 Quality coverage: 9.7 in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 8997: contig of 8997 bp in length  
 \* 8998 9097: gap of 100 bp  
 \* 9098 9921: contig of 824 bp in length  
 \* 9922 10021: gap of 100 bp  
 \* 10022 11236: contig of 1215 bp in length  
 \* 11237 11336: gap of 100 bp  
 \* 11337 12662: contig of 1326 bp in length  
 \* 12663 12762: gap of 100 bp  
 \* 12763 15814: contig of 3052 bp in length  
 \* 15815 15915: gap of 100 bp  
 \* 15915 27843: contig of 11929 bp in length  
 \* 27844 27943: gap of 100 bp  
 \* 27944 100820: contig of 72877 bp in length  
 \* 100821 100920: gap of 100 bp  
 \* 100921 172522: contig of 71602 bp in length  
 \* 172523 172623: gap of 100 bp  
 \* 172623 202198: contig of 29576 bp in length.

## FEATURES

source

1. 202198  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="np23-406024"  
 /clone\_lib="RPCI-23 Female Mouse BAC"

misc\_feature

1. 8997  
 /note="assembly\_fragment"  
 clone end:SP6  
 vector\_side:left"

misc\_feature

9098..9921  
 /note="assembly\_fragment"

misc\_feature

10022..11236  
 /note="assembly\_fragment"

misc\_feature

11337..12662  
 /note="assembly\_fragment"

misc\_feature

12763..15814  
 /note="assembly\_fragment"

misc\_feature

15915..172623  
 /note="assembly\_fragment"

misc\_feature

27944..100820  
 /note="assembly\_fragment"

misc\_feature

100921..172522  
 /note="assembly\_fragment"

misc\_feature

172623..202198  
 /note="assembly\_fragment"

misc\_feature

clone end:T7  
 vector\_side:right"

## RESULT 16

AC122549

LOCUS

DEFINITION

AC122549

AC122549

HTG: HTGS PHASE1; HTGS\_DRAFT.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC122549 202198 bp DNA linear HTG 13-MAR-2003  
 Mus musculus clone RP23-406024, WORKING DRAFT SEQUENCES, 9 unordered  
 pieces.

AC122549

AC122549

HTG: HTGS PHASE1; HTGS\_DRAFT.

KEYWORDS

SOURCE

ORGANISM

Mus musculus

(house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

1 (bases 1 to 202198)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,

Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K.,

Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,

Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,

Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,

Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J., Nguyen, C.,

Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N.,

Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,

Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,

Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,

Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,

Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,

Young, G., Zainoun, J., Zembek, U., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 202198)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,

Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,

Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,

Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K.,

Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,

Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,

Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,

Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,

Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,

Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,

Meldrum, J., Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J.,

Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,

O'Neil, D., Oliver, J., Peterson, K., Phukhang, P., Pierre, N.,

Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,

Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P.,

Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,

Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,

Vassiliev, H., Venkataran, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,

Wyman, D., Young, G., Zainoun, J., Zembek, U., Zimmer, A. and Zody, M.

Direct Submission

Submitted (13-MAR-2003) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 13, 2003 this sequence version replaced gi:28261613.

All repeats were identified using RepeatMasker:

Snit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center



misc\_feature  
/note="assembly fragment"  
202882..204171-  
/note="assembly\_fragment  
clone\_end:17  
vector\_side:right"

## ORIGIN

Query Match 6.0%; Score 28; DB 2; Length 204171;  
Best Local Similarity 100.0%; Pred. No. 0.00031;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAAA 81  
|||||  
DB 53830 GCTGCTGCTGCTGCTGCTGCTTAAA 53857

## RESULT 18

AY023222  
LOCUS AY023222 227 bp DNA linear PLN 07-FEB-2001  
DEFINITION Oryza sativa microsatellite MRS5547 containing (GCT)X9, closest to  
marker L825, genomic sequence.  
ACCESSION AY023222  
VERSION AY023222.1 GI:12706438  
KEYWORDS

## SOURCE

Oryza sativa  
Oryza sativa

ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Erihartoideae; Oryzeae; Oryza.

## REFERENCE

1 (bases 1 to 227)  
Tao, N., Barbazuk, W.B., Liu, J., Wu, K. and Barry, G.F.

AUTHORS  
TITLE Simple sequence repeats from Monsanto rice genomic sequences  
JOURNAL Unpublished

## REFERENCE

2 (bases 1 to 227)  
Tao, N., Barbazuk, W.B., Liu, J., Wu, K. and Barry, G.F.

AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh

Blvd., Creve Coeur, MO 63167, USA  
Derived from rice genomic sequences generated from the Monsanto  
Rice Genome Sequencing project. Please see  
<http://www.rice-research.org> for more information. The sequence  
data were produced primarily in the laboratories of Dr. Leroy Hood  
at the University of Washington in Seattle.

## FEATURES

source  
1..227  
/organism="Oryza sativa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4530"  
repeat\_region  
1..227  
/note="microsatellite MRS5547"  
/rpt\_type=tandem  
/rpt\_unit="gct"

## ORIGIN

Query Match 5.8%; Score 27; DB 8; Length 227;  
Best Local Similarity 100.0%; Pred. No. 0.00094;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77  
|||||

DB 98 AACGCTGCTGCTGCTGCTGCTGCTGCT 124  
|||||

## RESULT 19

G09506  
LOCUS G09506 373 bp DNA linear STS 14-AUG-1995  
DEFINITION human STS CHLC.GCT4G01.P10966 clone GCT4G01, sequence tagged site.  
ACCESSION G09506  
VERSION G09506.1 GI:941385

KEYWORDS STS; STS sequence; primer; sequence tagged site.

## SOURCE

Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

AUTHORS Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buetow, K.H.  
TITLE Cooperative Human Linkage Center  
JOURNAL Unpublished (1995)  
COMMENT Synonyms: GCT4G01, CHLC.GCT4G01.T10965  
Contact: Dr. Jeffrey C. Murray

## COMMENT

UofI  
The University of Iowa  
Department of Pediatrics, Iowa City, IA 52242, USA  
Tel: (319) 356-3508  
Fax: (319) 356-3347  
Email: jeff-murray@uiowa.edu

Primer A: TCGCGGTAAAAGATGAATG

Primer B: ACTCTAGGAATGAGACATTCGG

STS size: 106

PCR Profile:

denature: 30 seconds at 94 degrees C  
annealing: 75 seconds at 55 degrees C  
extension: 15 seconds at 72 degrees C  
PCR cycles: 27  
extension: 6 minutes at 72 degrees C

## Protocol:

Template: 30ng genomic DNA  
Primer: each 1.5 pmole  
dNTPs: each 200 uM  
Taq Polymerase: 0.3 units  
Total Vol: 10 ul

## Buffer:

MgCl2: 1.5mM  
KCl: 50mM  
Tris: 10mM  
pH: 8.3.

## FEATURES

source  
1..373  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
STS  
primer\_bind  
12..117  
primer\_bind  
12..31  
complement(96..117)  
ORIGIN

Query Match 5.8%; Score 27; DB 11; Length 373;  
Best Local Similarity 100.0%; Pred. No. 0.00096;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 CGCTGCTGCTGCTGCTGCTGCTGCTTA 79  
|||||

DB 69 CGCTGCTGCTGCTGCTGCTGCTGCTTA 95  
|||||

## RESULT 20

AB028183/c  
LOCUS AB028183 1264 bp mRNA linear PLN 22-JAN-2000  
DEFINITION Oryza sativa mRNA for OsNAC4 protein, complete cds.  
ACCESSION AB028183  
VERSION AB028183.1 GI:6730937

KEYWORDS OsNAC4; OsNAC4 protein.

## SOURCE

Oryza sativa

## ORGANISM

Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Erihartoideae; Oryzeae; Oryza.

## REFERENCE

1 (sites)  
Kikuchi, K., Ueguchi-Tanaka, M., Yoshida, K.T., Nagato, Y.,  
Matsusaka, M. and Hirano, H.Y.  
TITLE Molecular analysis of the NAC gene family in rice  
JOURNAL Mol. Gen. Genet. 262 (6), 1047-1051 (2000)  
MEDLINE 20123461  
PubMed 10660065

REFERENCE 2 (bases 1 to 1264)  
 AUTHORS Hirano, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-MAY-1999) Hirovuki Hirano, Univ Tokyo, Graduate School of Agricultural and Life Sciences, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan (E-mail:ahirano@mail.ecc.u-tokyo.ac.jp, Tel:+81-3-5841-5065, Fax:+81-3-5841-5063)  
 FEATURES  
 source  
 1..1264  
 /organism="Oryza sativa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4530"  
 gene  
 1..1264  
 /gene="OsNAC4"  
 CDS  
 100..1050  
 /gene="OsNAC4"  
 /codon\_start=1  
 /product="OsNAC4 protein"  
 /protein\_id="EAA89798.1"  
 /db\_xref="GI:6730938"  
 /translation="MAAAGVSGRRDAEALNPPGFRFHTDDELVHLCRKYARQ PLPPIIAEVDLDPDLPKALFGKKEWYFFPRDRKYPNGSRNRAAGRYWKA TGADKVPAPGKSGARTVGIKKALVYSGKAPGVKTDWIMHEYRLADRAPGKKGSG KLDEWLICRLYNKNWKKVLEOODVASVAAPRHHQNGEWMDEAAADPMDSDF QTHSDSDNAGLCRHGCGCGGFGDVAPRANGFVTVKEDNDFGLNFDLOPPYMW NLQHMQMWNPNAPAGDHGGYLSISSPQMKMMQTLPPF"  
 polyA\_site  
 1239  
 /gene="OsNAC4"  
 /note="25 a nucleotides"  
 ORIGIN  
 Query Match 5.8%; Score 27; DB 8; Length 1264;  
 Best Local Similarity 100.0%; Pred. No. 0.001;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 51 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77  
 Db 47 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 21  
 RESULT 21  
 AK073848/c  
 LOCUS AK073848 1283 bp mRNA linear PLN 24-JUL-2003  
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J033069M07, full insert sequence.  
 ACCESSION AK073848  
 VERSION AK073848.1 GI:32983871  
 KEYWORDS FLI CDNA; CAP trapper.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
 REFERENCE 1  
 The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team: Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki Y., Ishikawa M., Yamada H., Ooka H., Hotta I., Kojima K., Namiki T., Ohneda E., Yahagi W., Suzuki K., Li C., Ohtsuki K., Shishiki T., Otsuki K., Shishiki T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Ootomo Y., Murakami K., Iida Y., Sugano S., Fujimura T., Suzuki Y., Tsunoda Y., Kuroaki T., Kodama T., Masuda H., Kobayashi M., Xie Q., Lu M., Nariawa R., Sugiyama A., Mizuno K., Yokomizo S., Niikura J., Ikeda R., Ishibiki J., Kawamata M., Yoshimura A., Miura J., Kusumegi T., Oka M., Ryu R., Ueda M., Matsubara K., RIKEN, Kawai J., Carninci P., Adachi J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Hayashida K., Hayatsu N., Sano H., Sasaki D., Sato K., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagami-Takeda Y., Tagawa A., Takahashi F., Takaku-Akahira S., Tanaka T., Tomaru A., Toya T., Waki K., Yasunishi A. and Hayashizaki Y.  
 TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from Japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)  
 MEDLINE 22752273  
 PUBMED 12869764  
 REFERENCE 2 (bases 1 to 1283)  
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, P., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Inamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Nariawa, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)  
 COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.  
 URL : http://cdna01.dna.affrc.go.jp/cdna/  
 NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.  
 FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Nariawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.  
 Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, P., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.  
 FEATURES  
 Location/Qualifiers  
 1..1283  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="J033069M07"  
 ORIGIN  
 Query Match 5.8%; Score 27; DB 8; Length 1283;  
 Best Local Similarity 100.0%; Pred. No. 0.001;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 51 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77



```

YSTTFLFLINSILOCFNNKTALEILNFKLKNNIHNTITTFLELAYKDDPNSVWIMEQIK
QFKSESAIFDSMIKTNDTNSLPTKDIARIESEPLECLENCLLKKADSPVERPINE
IIQSLWKLDISQPSYKILLKLINLSLFLVLDISFOVSTNPSDFEILEAENVVYVF
QDSVKNLLDSIQDEIARRATEIDDLNAKISHLNKLNLENHDKHLLAKLDESSES
LISUKTEIENLKLQKATKRLQDITTHQRKYDPPSLASSNLSIAGSIKNNSHGN
IIIFONLAKKQOQKISLPRKSTSLKSKRVTSLSYLTIDANNENESSESKSDS
LFQSTSTINFIENSMKNITNMQVNSLSILEFNSLGTQPNYQSSPVLGSVSS
PKLFPRLSDSLDNGIQLVPEVKLPOLPPPPPPPPPLPOSILLTEARAKPDGVSCI
AAPAPPLPDLTKTKTCGAVPPPPPPPLPESLMMNKGPSNHDILVTPAPPPLNGLLS
SSVSINPTTDLKPPTEKELKQIHWKVEDIKOTLWEDTFQOETIKSLQDIDGIFS
QIEDIFKMSPTKIANKRNABESSLNNKSGSNELUKKISFLSRDLAQOFGINLHMF
SLSQDMEFVMVLCNDNDIVONVNLKFCFKEELVNTPKMLANKYEPSQKDGKAYS
DLQADRIIFELCINLRFYNNARKSLTSTYERDYDILFKLQIDDDASHLNSP
KFKSLMIIITEIGNHMKRIYKIGIKLSLTKLAFVRSSIDQNVSLFHLIEKVIKIKYP
DIYGVLDLKNIEDILSEHVESECHFEKKIDYAVTOFOVSKLSKEENLAPROQI
IKVVKFINAKTSEILIGOCKLTLDLANKMKYGEDPKDKSKNEKNEFFQDPIEFILK
MFKKAKENLEKEMERIVYQRKSLDMDRTSSNKKSGNSGDEGKNEFFQDPIEFILK
LREVKKDPBFLRRKRSKULNEIALINVEHGDVTKRKDBDHVLLERTHAMLNDIQNI"
7230. .7844
/notes="unnamed protein product; orf, 204aa, CAI=0.12"
/codon start=1
/protein_id="CAA86120.1"
/db_xref="GI:557765"
/db_xref="GOA:P40451"
/db_xref="SGD:S0001420"
/db_xref="SWISS-PROT:P40451"
/translations="MGNVAVGTAVGIPIAVGVIIALIPWCKLORRYKKEEIRDAL
EKWMEVAVSVVDGKAEVNSSSEASTNEKANQDLKPCQKTKAKAGTTPAYRQOL
NASMGTLRPKQSTATNVPIVPSGKYNKGWDEFSYFWYPLTILSRKETSLSRSAS
TSNLSSTENTALHEIKLDDPDEFTNTVNRKBEFDSLREH"
complement (7978. .8571)
/notes="unnamed protein product; orf, len 197, CAI: 0.19"
/codon start=1
/protein_id="CAA86121.1"
/db_xref="GI:557766"
/db_xref="GOA:P40452"
/db_xref="SGD:S0001419"
/db_xref="SWISS-PROT:P40452"
/translations="MMLRLVTKGLPKVTPSAKAVLVRSGLLHSFSTARSFNNSVAED
EAKVLKDNRLPIRELDPPTERRKRIAGLFLSVALGSALSILFNVKTESPII
SNTLYVIRSPATKILGSIEPDGIIIPWVYGLNSVKGRINITFIYKGDKNVGTGR
LVADRNDHEEFLLHWSVTAQOKLIDLAENKTEPI"
9120. .12335
/notes="unnamed protein product; orf, len: 1071, CAI: 0.13,
some similarity with ubiquitin carboxy terminal
hydrolases"
/codon start=1
/protein_id="CAA86122.1"
/db_xref="GI:557767"
/db_xref="GOA:P40453"
/db_xref="SGD:S0001418"
/db_xref="SWISS-PROT:P40453"
/translations="MLDDDKGTAMHPHTPTPEYSNELLRRVQDLYHEDIKHVYPOL
KLEKLDLLEHTPEYLPDYLDLSIHHDRPNDALTAETIGCYVYFLIIPQSLQOTRNS
YSITDCLKMYENENMTNVLVYKKEIGVLDSEVKGAGIEHRIITKRAFSVAD
LSQVASLSDTAAPQDGLKGTFTEDDASQSPVWTAPEENPQDLKALLPVPIPT
PAFREPKTSVPSRVLLEDVPSIYHEDDTSFASINPPREITADRSVTHRKDSYH
SYWVDGSLKEDNDLNFVNDGFIQSILDIKQSIITAPELFSILSNRVERKVLV
IDLIRPQRAGNIHIVAPLVVNDVNLMDKQNTPIYKDDIIEHLIKENENFINRKF
DIVYITDVTMTNFDYAFIFEVYLMLTQSQTEPLTPTVTLILGGYKWKTLHSIAQ
EXHISIEDLYRPSYKARLQEQOQQQPSQDSFSAKESSTKVPPSPWPKPDLPL
ILNRKPPPPPPVMTPEIPIPPPKIMVHSQVSSSRKPPIPAKOYHKEQLNSN
EIIRKQRQHQHQIQLQORAYNIPTIERSPNVYSLITGLRNLGNTCYINSMI
QCLFAKTRFTLFISSKYRSLQIRNSGSHYSPKLSNLSMLNKMVINGCSVPT
GFLKINQRLPDLKIPDDQDQOEFLMILLDRHLDLSLQOQVANDYNLLYNADAL
KYSNNEYKHFQKQVINGISPIDIFQGMENSLQCKRCQITNTSYTFVLSAIP
RSMKSLKLRCTEKKRLDCINLTFQSMENSLQCKRCQITNTSYTFVLSAIP
EFSIVKSKKKSRFFTLHTGTRHLDLFGDGTTEGNSNNNTTIFERSRSPFRM
LGSGKRSSTSTPFGTGNDSNSDYKDKLTITVKTNFVTLPKLILVHSRYVDL
TKNNNTVTVYPLKNTILKNDNTMKYKLPGVVNHGTLSLGHYTSVNLKDLHNVNIG
RSKMYVDFDEVVKADRKGSKDLKISSSDVVVLYERYVD"
complement (12788. .14737)
/genes="gut2"

```

## CDS

```

complement (12788. .14737)
/genes="gut2"
/notes="gut2, len: 649, CAI: 0.19, GPDA YEAST P32191
GLYCEROL-3-PHOSPHATE DEHYDROGENASE, frameshifts with
published sequence"
/codon start=1
/product="glycerol-3-phosphate dehydrogenase"
/protein_id="CAA86123.1"
/db_xref="GI:557768"
/db_xref="GOA:P32191"
/db_xref="SGD:S0001417"
/db_xref="SWISS-PROT:P32191"
/translations="MFSVTRRAAGAAAMATATGTYWMTSGDRPLVHNDSPYVQ
FTAPAPQVSRDLDLAKTHQFDVLIIGGATGTCALDAATRLNALVEKGDA
SGTSSKTXMIRGVVLYLEKAFWESKQADLVIEALNERKHLINAPLHCLTVLPII
PIYSWQVPIYNGCKFYDFAGSQNLKSYLLSKSATVEKAPMLTDTNLKASLVYHD
GSFNRLNATLAI TAVENGATVNVYEVOKLIKOPTSKVIGAEARDVETMLVPIIN
AKCVNATGPGSDAILQMDRNPGLPDSPLNDSKIKSTFNGLVANDPKMVPISGVH
IVLPSPYCPKMDLLDVRTSDGRVFFLPWQSKVLGATDIDPLKQVFPNMPTEADIQ
DILKELQHYIEFVKREDVLSAWGVRLVRDPTIPADGKGSATQGVVSRHFLFTS
DNGLITTAGGKWTYRQMAETVDKYVEGVGFHNLKPCHTRIDKLAGEEWTQYVAL"

```

## Query Match

```

Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 53 CGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTTA 79
|||||
Db 17842 CGTGTCTGCTGCTGCTGCTGCTGCTGCTTA 17816

```

## RESULT 24

```

AP004331
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
BAC clone: OSJNBa0085D07, complete sequence.
ACCESSION
AP004331 BA000010
VERSION
AP004331.2 GI:21203157
KEYWORDS
HTG.
SOURCE
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
1
Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
Katayose, Y., Wu, J., Nilmura, Y., Cheng, Z., Nagamura, Y.,
Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y.,
Idonuma, A., Iijima, M., Ikeda, M., Ikono, M., Itoh, S., Itoh, T.,
Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
Ohta, I., Ono, N., Sai, S., Sakai, K., Shibata, M., Shimokawa, T.,
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,
Yanagata, H., Yamane, H., Yoshiki, S., Yoshihara, K., Yukawa, K.,
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
Yano, M., Jiang, J. and Gojobori, T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)
22337376
MEDLINE
PUBMED
12447438

```

## REFERENCE

```

AUTHORS
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (14-NOV-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On May 24, 2002 this sequence version replaced gi:16930116.
COMMENT

```

## gene





```

TGADKPVAKGARTVGIKALVYSGKAPGVKTDIMHXYRLADADRACPGCKGSGQ
KLDWVLCRLYNKKNNEKYKLEQDDVAVAAAPRRHHNGVMRAAADIMSDSF
QTHSDIDNAGRHGGGGGGGDDVAPPNGMGTIVKEDNWTGLNFDELOPPYMM
NLHQMQMNGAPAPAGDDGGYQISSPGMMQWITLPPF"
join(6090) .6317,6421. .6572,8325. .8440,8715. .8803,
9502. .9589,9669. .9774,10773. .10866,11020. .11192,
11536. .11575,12734. .12875,13216. .13303,13676. .13730)
/gene="B1148D12.2"
/gene="B1148D12.3"
join(6090) .6317,6421. .6572,8325. .8440,8715. .8803,
9502. .9589,9669. .9774,10773. .10866,11020. .11192,
11536. .11575,12734. .12875,13216. .13303,13676. .13730)
/gene="B1148D12.2"
/note="contains EST"
hypothetical protein"
/codon_start=1
/protein_id="BAB64821.1"
/db_xref="GI:15528780"
/translaton="MRAACEGGMWTAAGLASRWGTSAAGSGGSLREELGNHANAVA
TSFLRHRSQLRWEQGAQKRGTVGNAPFSEASMEYAPSPSIALPSDPPPAASRV
ACRLYHRSQCPHGLEPFLPPARGHACTHALTGNPQPAVTGTYVLACNPFCDYC
RVASQNTVRMSNCEASTRTVVGSTPRQEDTQQTGTGLSPVTFEDRTGCAPNFDYC
PNGVAGGVGLMSVSGGRWGLEDRLLSFQAKAQIVVRACGLYNSTRQPLAAAT
ESQDDVTRPRRGEHVEHLGCYYGYGRENVTVLSTMHPRCPCLNSTRSSMIGS
SMSCGRGFQSRQIKKVDYSVTGDQLHVELAHWFHNRHRTAAAGGQSQSYVVA
ESQCRPNTAAGLHGRVFSRLVAYYTCNVCYLRCSLYVDCSRTPLPLRKQVH
LHLSGK"
complement(join(14359) .14600,14752. .14929,15013. .15084,
15363. .15458,15562. .15678,15760. .15916,16010. .16107,
16180. .16338,16409. .16567,16788. .16907,17208. .17325,
17456. .17586,17692. .17768,18098. .18209,18297. .18479,
18656. .18757,18857. .18958,19057. .19152,19721. .19798,
19892. .19984,20057. .20178,20275. .20356,20444. .20518,
20625. .20885,20970. .21094,21196. .21271,21368. .21517,
21599. .21678,21764. .21865,21947. .22054,23134. .23246,
23518. .23623,24383. .24487,24596. .24667,24690. .24847,
25307. .25415,25492. .25557,25644. .25780,26028. .26130,
26656. .26775,26947. .27060,27163. .27248,27328. .27410,
28523. .28587,28684. .28861,28937. .29016,30166. .30289,
30380. .30485,30639. .30801)
/gene="B1148D12.3"
complement(join(14359) .14600,14752. .14929,15013. .15084,
15363. .15458,15562. .15678,15760. .15916,16010. .16107,
16180. .16338,16409. .16567,16788. .16907,17208. .17325,
17456. .17586,17692. .17768,18098. .18209,18297. .18479,
18656. .18757,18857. .18958,19057. .19152,19721. .19798,
19892. .19984,20057. .20178,20275. .20356,20444. .20518,
20625. .20885,20970. .21094,21196. .21271,21368. .21517,
21599. .21678,21764. .21865,21947. .22054,23134. .23246,
23518. .23623,24383. .24487,24596. .24667,24690. .24847,
25307. .25415,25492. .25557,25644. .25780,26028. .26130,
26656. .26775,26947. .27060,27163. .27248,27328. .27410,
28523. .28587,28684. .28861,28937. .29016,30166. .30289,
30380. .30485,30639. .30801)
/note="contains EST C25950(c11258)"
/codon_start=1
/product="putative microtubule associated protein"
/protein_id="BAB64822.1"
/db_xref="GI:15528781"
/translaton="MSTDEKLKAEKLPWDLRLOHKNWKNVNDANIDLAALCDSIT
DPKDLRREGPLFKTVDASNPQKALDALLAFORAADADASYAKEVCDAIVAK
CLTRKPTKKAQAFLLWYLEAAEVLSEKAVKNVAKAVPAIDVFMQALSEF
GAKVPPKLLKMLPELDPDQNVRASSKGLTLELCRWIGKEPVKAILPEKRMDFW
KELEAELVSGIAPKPRKIRSOEKELEEVPEAAGTNNSEEAPEAPMEIDEVLI
VDVDDILTPLEKSGFVGKATKWSERDAVELTKLASTKTAQDFHEICRLKLL
ITDNLAVSVEAQAIGNLAKRLHRSFGNSRVLPLVLEKKEKPTTEALUSQTLQ
AMHSGCITLDDVEDVRVAVKNKPLVRLSLTNWAFCTISNKAIVLKHXEYPI
CMELNDTPEVDSFSLVTAIAKVMGKPLERLEKLDVVRKKLSMDIGASDIT
SGTEGSIQIDIIPEFGKEMLCRAVTVPVTLINDYELLARGRLVAFVLDIAITDVPW
DSSSMRASAASMLSGKPPQAVPATKSGPAKATKTKDGGPQSQKASAPVIEDYEV
SEMSLEFIEKLSLVKSETISQLKSTVWKELEASMLKQEVESITELDKSLELIV
LLCAVFGWSENQVLYIKMKHKNPKVLSEGLLWMVSAVEDFGISLNUKOTIDFCO
IGLOSSAAATRNATIKLIGVJHFKVGPDIKGFSLDKVPALLSTLDAEYKNEPFGTAS

```

```

APRTRVAAADVSSASGTSGLPREDISAKITPTLLKLNIGSPWKLRLSIEDAVSKI
VEAHKRIQPTGVLFETALRLYDSNNKLVNATLTIGTIGLASEGPAVEKSGKGL
ADVLKCLGDNKEMRECTIATLDLWAAALDKMVPYITVLTGPKTQKXDLFPW
LSXGASNDSPSEALPLPKPSASSLMDKSEVRKAASFNNEILIKICQDVAANLKO
LPSPTLAIVAEELIKLSTVHGFSDSVKMTTSMLSKAGSKNNKHGPNDRGSVSKA
VSQRGFPARSVTMISDQSIQQAFLNIKDSNKEERERVLVRKFEEPRRRQIIDE
LKLELFFRSFVSLRLWNSDFRQIQDIELLOKALPSSRSEVLELJLILRLVFLRF
CESNTCLLKVLDFLPELFDVKDQOSYMLTEAAAIPLFCIMKESGNEVSGLLKNPSV
LIFQMNITYSLPKLLPYILEGRSKNNRTIECVDIITGYFMDHGTGNEVSGLLKNPSV
AALTAREDSIRKAALNTLATAYKNGLDVVRVYVGLSDAQRSLMDRFFWKAREMDK
RRSGPCDARAALRRSVRENGSDIAEGSGEAVSRMAGENFGYSDAMVPRQATATP
GDMREALDIVALGLPEQSGEKMVICHELTONVDEPSSVLDLKEADRLVSLAV
MTQIKRLAHVKEGLDNLITELLMLDERVFLMDDGSQLKALNVLKMLKIDNAE
RTSFFVLNLLRLDPSRNPSTPPESLAVKNOKFSDLVVKCLIKLTKVQSTIYEV
DLDRILQSTHIYLOELGMEIRRRAGADKPLRVKTVLHELKLGRTAIKGLHSMVP
IDRAEPQIITILAYIDLNLQTLAAARMLTPSGTMGTQHWGAGSNPNPSTHSTDLQKQ
ELAAFKGIQKOTCTIGLYELVITQLYPKVDIFAQLONASAFRTYIRDLAQVFK
NAAAGTPSSLESTPTTPIAPISPKFAPSVPHTKSNKTKDCNEDDAGDTHPFGT
LDALRRMSKIQAAAVGANFDGVQAPLPSFMNGNTLHGGTRLDADTQNTLIIPMDER
ALSGLQARMERLKSQMEPL"
complement(41340) .43701)
/gene="B1148D12.4"
complement(41340) .43701)
/note="contains EST"
probably inactive due to no termination codon in CDS
pseudogene
similar to Arabidopsis thaliana chromosome 2, At2g35630"
/pseudo
join(46995) .47010,47163. .47331,48588. .48659,48754. .48915,
49020. .49094,49172. .49243,49329. .49400,49485. .49553,
49972. .50106,50186. .50655,50730. .50888,50981. .51120,
51214. .51433,51545. .51780,51867. .52370)
/gene="B1148D12.5"
join(46995) .47010,47163. .47331,48588. .48659,48754. .48915,
49020. .49094,49172. .49243,49329. .49400,49485. .49553,
49972. .50106,50186. .50655,50730. .50888,50981. .51120,
51214. .51433,51545. .51780,51867. .52370)
/note="contains EST"
/codon_start=1
/product="putative receptor protein kinase"
/protein_id="BAB64823.1"
Query Match 5.8%; Score 27; DB 8; Length 132059;
Best Local Similarity 100.0%; Pred.No.0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
|||||
Db 2817 AACGCTGCTGCTGCTGCTGCTGCTGCT 2843
|||||
RESULT 27
AC103357 143558 bp DNA linear HTG 04-DEC-2003
LOCUS Mus musculus chromosome 14 clone RP24-201114 map 14, *** SEQUENCING
IN PROGRESS ***, 2 ordered pieces.
AC103357 GI:38678639
HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 143558)
Mus musculus chromosome 14, clone RP24-201114
Unpublished
2 (bases 1 to 143558)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,I., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,

```

	FEATURES	SOURCE
1	100% cotton	100%
2	100% cotton	100%
3	100% cotton	100%
4	100% cotton	100%
5	100% cotton	100%
6	100% cotton	100%
7	100% cotton	100%
8	100% cotton	100%
9	100% cotton	100%
10	100% cotton	100%
11	100% cotton	100%
12	100% cotton	100%
13	100% cotton	100%
14	100% cotton	100%
15	100% cotton	100%
16	100% cotton	100%
17	100% cotton	100%
18	100% cotton	100%
19	100% cotton	100%
20	100% cotton	100%
21	100% cotton	100%
22	100% cotton	100%
23	100% cotton	100%
24	100% cotton	100%
25	100% cotton	100%
26	100% cotton	100%
27	100% cotton	100%
28	100% cotton	100%
29	100% cotton	100%
30	100% cotton	100%
31	100% cotton	100%
32	100% cotton	100%
33	100% cotton	100%
34	100% cotton	100%
35	100% cotton	100%
36	100% cotton	100%
37	100% cotton	100%
38	100% cotton	100%
39	100% cotton	100%
40	100% cotton	100%
41	100% cotton	100%
42	100% cotton	100%
43	100% cotton	100%
44	100% cotton	100%
45	100% cotton	100%
46	100% cotton	100%
47	100% cotton	100%
48	100% cotton	100%
49	100% cotton	100%
50	100% cotton	100%
51	100% cotton	100%
52	100% cotton	100%
53	100% cotton	100%
54	100% cotton	100%
55	100% cotton	100%
56	100% cotton	100%
57	100% cotton	100%
58	100% cotton	100%
59	100% cotton	100%
60	100% cotton	100%
61	100% cotton	100%
62	100% cotton	100%
63	100% cotton	100%
64	100% cotton	100%
65	100% cotton	100%
66	100% cotton	100%
67	100% cotton	100%
68	100% cotton	100%
69	100% cotton	100%
70	100% cotton	100%
71	100% cotton	100%
72	100% cotton	100%
73	100% cotton	100%
74	100% cotton	100%
75	100% cotton	100%
76	100% cotton	100%
77	100% cotton	100%
78	100% cotton	100%
79	100% cotton	100%
80	100% cotton	100%
81	100% cotton	100%
82	100% cotton	100%
83	100% cotton	100%
84	100% cotton	100%
85	100% cotton	100%
86	100% cotton	100%
87	100% cotton	100%
88	100% cotton	100%
89	100% cotton	100%
90	100% cotton	100%
91	100% cotton	100%
92	100% cotton	100%
93	100% cotton	100%
94	100% cotton	100%
95	100% cotton	100%
96	100% cotton	100%
97	100% cotton	100%
98	100% cotton	100%
99	100% cotton	100%
100	100% cotton	100%

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 103889: contig of 103889 bp in length
* 103890 103989: gap of unknown length
* 103990 131370: contig of 27381 bp in length
* 103991 131470: gap of unknown length
* 131471 158608: contig of 27136 bp in length.
* Location/Qualifiers
* 1..158608

```

OK 73019, USA  
 ON Dec 9, 2003 this sequence version replaced gi:38502358.  
 ----- Genome Center  
 Center: Department of Chemistry And Biochemistry  
 The University of Oklahoma  
 Center code:UOKNOR

```

/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-1316"
/clone_lib="Medicago truncatula BAC library H2"

ORIGIN
Query Match          5.8%; Score 27; DB 2; Length 158606;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 TAACGCTGCTGCTGCTGCTGCTGCTGCTG 76
|||||
Db 84940 TAACGCTGCTGCTGCTGCTGCTGCTGCTG 84914

RESULT 29
AC010921/c          163466 bp DNA linear INV 14-MAR-2001
LOCUS
DEFINITION
Drosophila melanogaster, chromosome X, region 15B-15E, BAC clone
BACR15L12, complete sequence.
AC010921
AC010921.11 GI:13324755
VERSION
KEYWORDS
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 163466)
REFERENCE
Celnikier,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busan,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Matti,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskaas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome X, region 15B-15E
Unpublished
2 (bases 1 to 163466)
REFERENCE
Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champet,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskaas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (27-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 14, 2001 this sequence version replaced gi:7143399.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
Location/Qualifiers
1. .163466
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"

FEATURES
source
1. .163466
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"

```

```

/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="15B-15E"
/clone="BACR15L12 (D898)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"

ORIGIN
Query Match          5.8%; Score 27; DB 3; Length 163466;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGCTGCTGCTGCTGCTGCTGCTGCTGCTTA 79
|||||
Db 94716 CGCTGCTGCTGCTGCTGCTGCTGCTTA 94690

RESULT 30
AL954848
LOCUS
DEFINITION
Zebrafish DNA sequence from clone CH211-281G7, complete sequence.
AL954848
AC010921.11 GI:29165221
VERSION
KEYWORDS
HIG.
Dario rerio (zebrafish)
Dario rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 170178)
REFERENCE
Johnson,C.
Direct Submission
Submitted (23-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 23, 2003 this sequence version replaced gi:29125265.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, ENBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep Repeat names
beginning 'Dr' were identified by the Recon repeat discovery system
(Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr'
were identified by Rick Waterman (Stephen Johnson lab, WashU). For
further information see http://www.projects/D\_rerio/fishmask.shtml
CH211-281G7 is from a CHORI-211 BAC library
VECTOR: pTABAC2.1
Location/Qualifiers
1. .170178
/organism="Dario rerio"
/mol_type="genomic DNA"

```

FEATURES  
source

```

/db_xref="taxon:7955"
/clonelib="CH211-281G7"
/clonelib="CHORI-211"

ORIGIN
Query Match      5.8%; Score 27; DB 5; Length 170178;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 AGCTGCTGCTGCTGCTGCTGCTGCTGCTT 78
|||||
Db 36000 AGCTGCTGCTGCTGCTGCTGCTGCTT 36026

RESULT 31
AC012160 172069 bp DNA linear INV 13-DEC-2001
Drosophila melanogaster, chromosome X, region 15E-15E, BAC clone
BAC06G02, complete sequence.
AC012160
AC012160.6 GI:17646875
HTG
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 172069)
REFERENCE
AUTHORS
Celniker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Baunton,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferrera,S., Frise,B., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Hocky,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pacleb,J., Paragac,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,
Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheller,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome X, region 15E-15E
Unpublished (1998)
2 (bases 1 to 172069)
Celniker,S.E., Agbavani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snitz,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (21-OCT-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 13, 2001 this sequence version replaced gi:6466928.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
1. 172069
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"

FEATURES
source
AC012160 172069 bp DNA linear HTG 10-DEC-2003
Medicago truncatula clone mth2-6c16, WORKING DRAFT SEQUENCE, 15
unordered pieces.
AC137828
AC137828.15 GI:39652644
HTG; HTGS PHASE1; HTGS DRAFT.
Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 172332)
Shaul,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Medicago truncatula BAC Clone mth2-6c16
Unpublished
2 (bases 1 to 172332)
Shaul,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (03-DEC-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Dec 10, 2003 this sequence version replaced gi:39228232.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2136: contig of 2136 bp in length
2137 2236: gap of unknown length
2237 4244: contig of 2008 bp in length
4245 4344: gap of unknown length
4345 7875: contig of 3531 bp in length
7876 7975: gap of unknown length
7976 10848: contig of 2873 bp in length
10849 10949: gap of unknown length
10949 14384: contig of 3436 bp in length

```



```

/rpt_family="L1"
7388. .7448
/rpt_family="B2"
38315. .38427
/rpt_family="B2"
38707. .38775
/rpt_family="ERV1"
41034. .41175
/rpt_family="Alu"
42647. .43068
/rpt_family="L1"
43072. .44040
/rpt_family="L1"
44162. .44313
/rpt_family="B2"
44381. .44438
/rpt_family="ERV1"
44495. .44556
/rpt_family="L1"
44557. .44744
/rpt_family="B2"
complement(44666. .44738)

Query Match      5.8%; Score 27; DB 10; Length 177062;
Best Local Similarity 100.0%; Pred.No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 110056 AACGCTGCTGCTGCTGCTGCTGCT 110082

RESULT 34
AC102553/c
LOCUS AC102553 185722 bp DNA linear HTG 18-JUN-2003
DEFINITION Mus musculus clone RP23-159D6, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
ACCESSION AC102553
VERSION AC102553.3 GI:31980154
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 185722)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-159D6
REFERENCE 2 (bases 1 to 185722)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chararo,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge MA 02141, USA
JOURNAL

```

REFERENCE  
AUTHORS

3 (bases 1 to 185722)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,  
 Collymore, A., Cook, A., Cooke, P., Corum, B., DeArillano, K.,  
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, B., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (18-JUN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 18 2003 this sequence version replaced gi:22380811.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
JOURNAL

## COMMENT

ORIGIN  
 Query Match 5.8%; Score 27; DB 2; Length 185722;  
 Best Local Similarity 100.0%; Pred. No. 0.0012;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 51 AACCTGCTGCTGCTGCTGCTGCTGCTGCT 77  
 Db 153507 AACCTGCTGCTGCTGCTGCTGCTGCTGCT 153481

## RESULT 35

## AC098642/c

## LOCUS

## DEFINITION

## AC098642

## sequence

## AC098642

## VERSION

## AC098642.5

## KEYWORDS

## HTG.

## SOURCE

## Mus musculus

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## This sequence version replaced gi:18201769.

## On Jul 25, 2002 this sequence version replaced gi:18201769.

## This sequence was finished as follows unless otherwise noted: all

## regions were either double-stranded or sequenced with an alternate

## chemistry or covered by high quality data (i.e., phred quality &gt;=

## 30); an attempt was made to resolve all sequencing problems, such

## as compressions and repeats; all regions were covered by at least

## one plasmid subclone or more than one M13 subclone; and the

## source

## Location/Qualifiers

## 1. 185722



```

assembly was confirmed by restriction digest.
FEATURES
  source
    Location/Qualifiers
      1..188834
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="RP23-27007"
        /clone_lib="RPC1-23"
ORIGIN
  Query Match
  Best Local Similarity 100.0%; Pred. No. 0.0012;
  Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 79536 AACGCTGCTGCTGCTGCTGCTGCTGCTGCT 79510
RESULT 36
AC114715
LOCUS
DEFINITION
  Rattus norvegicus clone CH230-339N2, WORKING DRAFT SEQUENCE, 4
AC114715
ACCESSION
  AC114715.5 GI:25008972
VERSION
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
  Rattus norvegicus
SOURCE
  Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE
  1 (bases 1 to 190402)
  Murny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
  Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
  Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
  Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
  Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
  Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
  Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
  Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J.,
  Cleveland,C., Cockrell,R., Cox,C., Coyte,M., Cree,A., D'Souza,L.,
  Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
  Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
  Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
  Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
  Fernandez,S., Finley,M., Flaggs,N., Forbes,L., Foster,M., Foster,P.,
  Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
  Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
  Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
  Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,J.,
  Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
  Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
  Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
  Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
  Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
  Lorensheva,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
  Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
  Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
  Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
  Milosavljevic,A., Miner,K., Minja,E., Montemayor,J., Moore,S.,
  Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
  Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
  Nwaokelimeh,O., Okwuonon,G., Olarnngyagoon,A., Pal,S., Parks,K.,
  Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankuch,C.,
  Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
  Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
  Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
  Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
  Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
  Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
  Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,K., Sosa,J.,

```

Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Unpublished  
2 (bases 1 to 190402)  
Worley,K.C.

Submitted (11-MAR-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 190402)  
Rat Genome Sequencing Consortium.  
Direct Submission

Submitted (15-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 15, 2002 this sequence version replaced gi:23681968.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

Center: Genome Center  
Center: Baylor College of Medicine

Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu

Project Information  
Center project name: GTHG

Center Clone name: CH230-339N2

Assembly program: Phrap; version 0.990329

Consensus quality: 173476 bases at least Q40

Consensus quality: 175237 bases at least Q30

Consensus quality: 176250 bases at least Q20

Estimated insert size: 172780; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).

NOTE: This is a 'working draft' sequence. It currently  
consists of 4 contigs. The true order of the pieces

is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence.

as soon as it is available and the accession number will  
be preserved.

1 172097: contig of 172097 bp in length

172098 172197: gap of unknown length

172198 187288: contig of 15091 bp in length

187289 187389: gap of unknown length

187389 188702: contig of 1314 bp in length

188703 188802: gap of unknown length

188803 190402: contig of 1600 bp in length.

Location/Qualifiers  
1..190402  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"

FEATURES  
source

```

/db xref="taxon:10116"
/clone="CH230-339N2"
915 1432
/notes="clone_boundary
clone_end:sp6
site:
end_sequence:BZ165181"
133344..133527
/notes="clone_boundary
clone_end:T7
site:
end_sequence:BZ165180"
172198..173628
/notes="wgs_end_extension
clone_end:T7"

ORIGIN
Query Match          5.8%; Score 27; DB 2; Length 190402;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAA 80
|||||
Db 5116 GCTGCTGCTGCTGCTGCTGCTGCTTAA 51142

RESULT 37
AC132877
LOCUS
DEFINITION
Mus musculus clone RP24-216J21, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION
AC132877
VERSION
AC132877.3 GI:28927732
KEYWORDS
HTG: HTGS PHASE1; HTGS DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 192539)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-216J21
Unpublished
2 (bases 1 to 192539)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Willson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,N.
Direct Submission
Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192539)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cooke,A., Cooke,P., Corrum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

```

```

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Willson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2003 this sequence version replaced gi:28412043.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27122
Center clone name: 216_J21
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191804 bases at least Q40
Consensus quality: 191385 bases at least Q30
Consensus quality: 191502 bases at least Q20
Insert size: 191939; sum-of-contigs
Quality coverage: 9.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7935: contig of 7935 bp in length
* 7936 8035: gap of 100 bp
* 8036 10717: contig of 2682 bp in length
* 10718 10817: gap of 100 bp
* 10818 21133: contig of 10316 bp in length
* 21134 21233: gap of 100 bp
* 21234 37038: contig of 15805 bp in length
* 37039 37138: gap of 100 bp
* 37139 73158: contig of 36020 bp in length
* 73159 73258: gap of 100 bp
* 73259 134748: contig of 61490 bp in length
* 134749 134848: gap of 100 bp
* 134849 192539: contig of 57691 bp in length.
*
* Location/Qualifiers
* 1..192539
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP24-216J21"
* /clone_lib="RPC1-24 Male Mouse BAC"
* 1..7935
* /note="assembly_fragment
* clone end:SP6
* vector side:left"
* 8035..10717
* /note="assembly_fragment"
* 10818..21133
* /note="assembly_fragment"
*
FEATURES
source
misc_feature
misc_feature
misc_feature

```

```

misc_feature 21234..37038
              /note="assembly_fragment"
misc_feature 37139..73158
              /note="assembly_fragment"
misc_feature 73259..134748
              /note="assembly_fragment"
misc_feature 134849..192539
              /note="assembly_fragment"
              clone_end:17
              vector_side:right"

ORIGIN
Query Match      5.8%; Score 27; DB 2; Length 192539;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
      |||||
DB 93242 AACGCTGCTGCTGCTGCTGCTGCTGCT 93268

RESULT 38
AC118589
LOCUS AC118589 192629 bp DNA linear HTG 11-MAR-2003
DEFINITION Mus musculus clone RP23-198G19, WORKING DRAFT SEQUENCE, 6 unordered
          pieces.
AC118589
VERSION AC118589.3 GI:2891232
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
          1 (bases 1 to 192629)
REFERENCE 1 (bases 1 to 192629)
          Birren,B.; Nusbaum,C. and Lander,E.
          Mus musculus, clone RP23-198G19
          Unpublished
REFERENCE 2 (bases 1 to 192629)
          Birren,B.; Linton,L.; Nusbaum,C.; Lander,E.; Ali,A.; Allen,N.;
          Anderson,S.; Barna,N.; Bastien,V.; Bloom,T.; Boguslavsky,L.;
          Boukhalter,B.; Brown,A.; Camarata,J.; Campopiano,A.; Chang,J.;
          Chazaro,B.; Choepel,Y.; Colangelo,M.; Collins,S.; Collymore,A.;
          Cook,A.; Cooke,P.; DeArellano,K.; Dewar,K.; Diaz,J.S.; Dodge,S.;
          Faro,S.; Ferreira,P.; FitzHugh,W.; Gage,D.; Galagan,J.; Gardyna,S.;
          Ginde,S.; Gord,S.; Goyette,M.; Graham,L.; Grand-Pierre,N.;
          Hagos,B.; Horton,L.; Hulme,W.; Iliev,I.; Johnson,R.; Jones,C.;
          Kamat,A.; Karatas,A.; Kellis,C.; LaRocque,K.; Lamazares,R.;
          Landers,T.; Lehoczy,J.; Levine,R.; Lindblad-Toh,K.; Liu,G.;
          MacLean,C.; MacDonald,P.; Major,J.; Marquis,N.; Matthews,C.;
          McCarthy,M.; McEwan,P.; McKernan,K.; Meldrum,J.; Meneus,L.;
          Mihova,T.; Mlenga,V.; Murphy,T.; Naylor,J.; Nguyen,C.; Nicol,R.;
          Norbu,C.; Norman,C.H.; O'Connor,T.; O'Donnell,P.; O'Neil,D.;
          Oliver,J.; Peterson,K.; Phunkhang,P.; Pierre,N.; Pollara,V.;
          Raymond,C.; Retta,R.; Rieback,M.; Riley,R.; Rise,C.; Rogov,P.;
          Roman,J.; Rosetti,M.; Roy,A.; Santos,R.; Schauer,S.; Schuback,R.;
          Seaman,S.; Severy,P.; Spencer,B.; Stange-Thomann,N.; Stojanovic,N.;
          Strauss,N.; Subramanian,A.; Talamas,J.; Tesfaye,S.; Theodore,J.;
          Topham,K.; Travers,M.; Travis,N.; Trigglio,J.; Vassiliev,H.;
          Viel,R.; Vo,A.; Wilson,B.; Wu,X.; Wyman,D.; Ye,W.J.; Young,G.;
          Zainoun,J.; Zembek,L.; Zimmer,A. and Zody,M.
          Direct Submission
          Submitted (19-APR-2002) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
          3 (bases 1 to 192629)
REFERENCE 1
AUTHORS Birren,B.; Nusbaum,C.; Lander,E.; Abouelleil,A.; Allen,N.;
          Anderson,S.; Arachchi,H.M.; Barna,N.; Bastien,V.; Bloom,T.;
          Boguslavsky,L.; Boukhalter,B.; Camarata,J.; Chang,J.; Choepel,Y.;
          Collymore,A.; Cook,A.; Cooke,P.; Corum,B.; DeArellano,K.;
          Diaz,J.S.; Dodge,S.; Dooley,K.; Dorris,L.; Erickson,J.; Faro,S.;
          Ferreira,P.; FitzGerald,M.; Gage,D.; Galagan,J.; Gardyna,S.;
          Graham,L.; Grand-Pierre,N.; Hafez,N.; Hagopian,D.; Hagos,B.;
          Hall,J.; Horton,L.; Hulme,W.; Iliev,I.; Johnson,R.; Jones,C.;
          Kamat,A.; Karatas,A.; Kellis,C.; Landers,T.; Levine,R.;

```

```

Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 11, 2003 this sequence version replaced gi:24850527.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17611
Center clone name: 198 G.19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191046 bases at least Q40
Consensus quality: 191531 bases at least Q30
Consensus quality: 191698 bases at least Q20
Insert size: 225000; agarose-fp
Insert size: 192129; sum-of-contigs
Quality coverage: 8.3 in Q20 bases; agarose-fp
Quality coverage: 9.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 117690: contig of 117690 bp in length
* 117691 117790: gap of 100 bp
* 117791 118991: contig of 1201 bp in length
* 118992 119091: gap of 100 bp
* 119092 129620: contig of 10529 bp in length
* 129621 129720: gap of 100 bp
* 129721 147127: contig of 17407 bp in length
* 147128 147228: gap of 100 bp
* 147228 186319: contig of 39092 bp in length
* 186320 186419: gap of 100 bp
* 186420 192629: contig of 6210 bp in length.
FEATURES
          Location/Qualifiers
          source            1..192629
                           /organism="Mus musculus"
                           /mol_type="genomic DNA"
                           /db_xref="taxon:10090"
                           /clone="RP23-198G19"
                           /clone_lib="RPCT-23 Female Mouse BAC"
          misc_feature     1..117690
                           /note="assembly_fragment"
                           clone_end:SP6
                           vector_side:left"
          misc_feature     117791..118991
                           /note="assembly_fragment"
          misc_feature     119092..129620
                           /note="assembly_fragment"
          misc_feature     129721..147127

```

/note="assembly fragment"  
147228..186319  
/note="assembly fragment"  
186420..192629  
/note="assembly fragment"  
clone\_end:T7  
vector\_side:right"

/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone="RP23-273G23"  
/clone\_lib="RPCI-23"

## ORIGIN

Query Match 5.8%; Score 27; DB 2; Length 192629;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 CGTGTCTGCTGCTGCTGCTGCTTA 79  
|||||  
Db 144931 CGTGTCTGCTGCTGCTGCTGCTTA 144957

RESULT 39  
AL929249/c  
LOCUS  
DEFINITION  
AL929249 Mouse DNA sequence from clone RP23-273G23 on chromosome 2, complete sequence.  
ACCESSION  
AL929249 GI:31043747  
KEYWORDS  
HTG.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 194209)  
Direct Submission  
Submitted (21-MAY-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On May 22, 2003 this sequence version replaced gi:30842758.  
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-273G23 is from the RPCI-23 Mouse BAC library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

FEATURES  
source  
1..194209  
/organism="Mus musculus"

## ORIGIN

Query Match 5.8%; Score 27; DB 10; Length 194209;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 AACGCTGCTGCTGCTGCTGCTGCT 77  
|||||  
Db 118602 AACGCTGCTGCTGCTGCTGCT 118576

RESULT 40  
AL845432/c  
LOCUS  
DEFINITION  
AL845432 Mouse DNA sequence from clone RP23-332H19 on chromosome 2, complete sequence.  
ACCESSION  
AL845432 GI:25955789  
KEYWORDS  
HTG.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 195835)  
Sycamore, N.  
Direct Submission  
Submitted (30-NOV-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
On Nov 29, 2002 this sequence version replaced gi:25137036.  
Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
-----  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-332H19 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6.

FEATURES  
source  
1..195835  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"

```

/chromosome="2"
/clone="RP23-332H19"
/clone_lib="RPCI-23"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 195835;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 38844 AACGCTGCTGCTGCTGCTGCTGCTGCT 38818

RESULT 41
AC124555
LOCUS
DEFINITION Mus musculus chromosome 14 clone RP23-245K17, complete sequence.
ACCESSION AC124555
VERSION AC124555.4 GI:38229416
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Wilson, R.K.
The sequence of Mus musculus clone
Unpublished
REFERENCE 2 (bases 1 to 210174)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 210174)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 210174)
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Nov 8, 2003 this sequence version replaced gi:31880269.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M BA0245K17
----- Location/Qualifiers -----
1. 210174
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/clone="RP23-245K17"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 210174;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAA 80
Db 206196 GCTGCTGCTGCTGCTGCTGCTGCTTAA 206222

RESULT 42

```

```

AL603682
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-235J5 on chromosome 11, complete
sequence.
ACCESSION AL603682
VERSION AL603682.36 GI:24394858
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211680)
Hopkins, B.
Direct Submission
Submitted (23-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:21952987.
----- Genome Center -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-235J5 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VBCTOR: pBACe3.6.

FEATURES
source
1. 211680
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-235J5"
/clone_lib="RPCI-23"

ORIGIN
Query Match          5.8%; Score 27; DB 10; Length 211680;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77
Db 25943 AACGCTGCTGCTGCTGCTGCTGCTGCT 25969

RESULT 43
AL604024
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-421F23 on chromosome 11,
complete sequence.
ACCESSION AL604024
VERSION AL604024.7 GI:21211848
KEYWORDS HTG.

```

SOURCE	Mus musculus (house mouse)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 217278)
AUTHORS	Whitehead, S.
JOURNAL	Direct Submission
TITLE	Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT	On May 25, 2002 this sequence version replaced gi:18477353. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <a href="http://www.sanger.ac.uk/projects/C_elegans/wormpep">http://www.sanger.ac.uk/projects/C_elegans/wormpep</a> RP23-421F23 is from the RPCI-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a> VECTOR: pBACe3.6.
FEATURES	<p>1..217278</p> <p>Location/Qualifiers</p> <p>/organism="Mus musculus"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:10090"</p> <p>/chromosome="11"</p> <p>/clone="RP23-421F23"</p> <p>/clone_lib="RPCI-23"</p>
ORIGIN	
Query Match	5.8%; Score 27; DB 10; Length 217278;
Best Local Similarity	100.0%; Pred. No. 0.0012;
Matches	27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	51 AACGCTGCTGCTGCTGCTGCTGCT 77
Db	58931 AACGCTGCTGCTGCTGCTGCTGCT 58957
RESULT 44	
AC116661/c	
LOCUS	AC116661 237094 bp DNA linear HTG 13-SEP-2002
DEFINITION	Mus musculus chromosome UNK clone RP23-11601, WORKING DRAFT SEQUENCE, 16 unordered pieces.
ACCESSION	AC116661
VERSION	AC116661.3 GI:22830524
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	McPherson, J.D. and Waterston, R.H.
JOURNAL	1 (bases 1 to 237094)
REFERENCE	The sequence of Mus musculus clone Unpublished
AUTHORS	2 (bases 1 to 237094)
TITLE	McPherson, J.D. and Waterston, R.H.
JOURNAL	Direct Submission
	Submitted (31-MAR-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA



```

Assembly program: Atlas 3.0;
Consensus quality: 203099 bases at least Q40
Consensus quality: 207416 bases at least Q30
Consensus quality: 210436 bases at least Q20
Estimated insert size: 213682; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 63260: contig of 63260 bp in length
* 63261 63360: gap of unknown length
* 63361 235896: contig of 172536 bp in length
* 235897 235996: gap of unknown length
* 235997 237332: contig of 1336 bp in length.
*
FEATURES             Location/Qualifiers
     source            1..237332
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-227F3"
     misc_feature      654..1459
                     /note="clone boundary"
                     clone_end:Sp6
                     site:EcORI
                     end_sequence:B2104009
     misc_feature      8193..18812
                     /note="clone boundary"
                     clone_end:T7
                     site:EcORI
                     end_sequence:B2104008
     misc_feature      63361..65231
                     /note="wgs end_extension"
                     clone_end:T7"

ORIGIN
Query Match      5.8%; Score 27; DB 2; Length 237332;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTTAA 80
Db 5161 GCTGCTGCTGCTGCTGCTGCTGCTTAA 5135

RESULT 46
AC099207/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-132D18, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC099207
AC099207.6 GI:30521989
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 240501)
Muzny, D., Marie, Metzker, M., Lee, A., Brannan, S., Adams, C., Alder, J.,
Allen, C., Allent, H., Altschuler, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

```

```

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Andra, C., Dederich, D.,
DeGado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Drapper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geier, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Guarinate, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenschewski, L., Louis, H., Lozano, R., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M., McNeill, T., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaakemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindester, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.,
Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C., Smajls, D.,
Sneed, A., Sodergren, E., Song, X., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D., Holt, R., Smith, H., O.,
Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 240501)
Worley, K.C.
Direct Submission
Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 240501)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25095914.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
-----
Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/

```



Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center project name: GHRS

Center clone name: CH230-132D18

----- Summary Statistics -----

Assembly program: Atlas 3.0;

Consensus quality: 227593 bases at least Q40

Consensus quality: 229405 bases at least Q30

Consensus quality: 230440 bases at least Q20

Estimated insert size: 236821; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 238878: contig of 238878 bp in length

\* 238879 238978: gap of unknown length

\* 238979 240501: contig of 1523 bp in length.

FEATURES

source

1. 240501

Location/Qualifiers

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-132D18"

1. 1103

misc\_feature

/note="wgs\_contig"

ORIGIN

Query Match 5.8%; Score 27; DB 2; Length 240501;

Best Local Similarity 100.0%; Pred. No. 0.0012;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCT 77

DB 47005 AACGCTGCTGCTGCTGCTGCTGCT 46979

RESULT 47

AC129456/c

LOCUS

AC129456 Rattus norvegicus clone CH230-9K8, WORKING DRAFT SEQUENCE.

DEFINITION

AC129456

ACCESSION

AC129456.5 GI:24940997

VERSION

HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 240561)

Muzzy, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Devilla, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, J., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M.,

Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., King, Z., Khan, Z., King, B., Kovar, C.,

Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenshuwa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokeme, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.,

Plopper, F., Poindexter, A., Popovic, D., Reeves, K., Regier, M. A., Reigh, R.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Richards, S., Riggs, F.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajd, D.,

Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Taylor, C.,

Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams, G., Willson, R., Wleczyk, R., Woodden, H., Worley, K.,

Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,

Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 240561)

Worley, K. C.

Direct Submission

Submitted (30-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 240561)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:23096544.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole

genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature

table.

----- Genome Center -----

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----

Center project name: GHRS

Center clone name: CH230-9K8

----- Summary Statistics -----

Assembly program: Phrap; version 0.990329

Consensus quality: 212559 bases at least Q40

Consensus quality: 214527 bases at least Q30

Consensus quality: 215680 bases at least Q20

```

-----
Estimated insert size: 218439; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 1 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
  the accession number will be preserved.
* the accession number will be preserved.
  1 240561: contig of 240561 bp in length.
  Location/Qualifiers
    1. 240561
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clones="CH230-9K3"
    1. 1414
      /note="wgs_contig"
      238534..240561
      /note="wgs_contig"

misc_feature
misc_feature

ORIGIN
Query Match          5.8%; Score 27; DB 2; Length 240561;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTGCTAA 80
Db 136293 GCTGCTGCTGCTGCTGCTGCTGCTAA 136267

RESULT 48
AC093480
LOCUS AC093480.3 GI:39752742
DEFINITION Mus musculus chromosome 13 clone RP23-61L24 map 13, HTG 12-DEC-2003
IN PROGRESS ***, 4 unordered pieces.
ACCESSION AC093480
VERSION AC093480.3
KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Birren,B., Nusbaum,C. and Lander,E.
1 (bases 1 to 242978)
Mus musculus chromosome 13, clone RP23-61L24
Unpublished
2 (bases 1 to 242978)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzGerald,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,

```

```

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-AUG-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 242978)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J., Choepe,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-DEC-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 12, 2003 this sequence version replaced gi:2325292.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L14395
Center clone name: 61_L_24
-----
* NOTE: This is a 'working draft' sequence. It currently
  consists of 4 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
  1 46051: contig of 46051 bp in length
  46052 46151: gap of 100 bp
  46152 48744: contig of 2593 bp in length
  48745 48844: gap of 100 bp
  48845 168635: contig of 119791 bp in length
  168636 168735: gap of 100 bp
  168736 242978: contig of 74243 bp in length.
  Location/Qualifiers
    1. 242978
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /chromosome="13"
      /map="13"
      /clone="RP23-61L24"
      /clone_lib="RPCI-23 Female Mouse BAC"

FEATURES
source
ORIGIN
Query Match          5.8%; Score 27; DB 2; Length 242978;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77

```

REFERENCE	AUTHORS	TITLE	JOURNAL
3	(bases 1 to 250546)	Rat Genome Sequencing Consortium.	
COMMENT			
<p>of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA</p> <p>Submitted (09-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA</p> <p>On May 9, 2003 this sequence version replaced gi:24818284. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.</p> <p>----- Genome Center  Center: Baylor College of Medicine  Center code: BCM  Web site: http://www.hgsc.bcm.tmc.edu/  Contact: hgsc-help@bcm.tmc.edu  ----- Project Information  Center project name: GBXY  Center clone name: CH230-6D11  ----- Summary Statistics  Assembly program: Atlas  Consensus quality: 223239 bases at least Q40  Consensus quality: 226328 bases at least Q30  Consensus quality: 228394 bases at least Q20  Estimated insert size: 233947: sum-of-contigs estimation  Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  -----</p> <p>* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).  * NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p> <p>1 249334: contig of 249334 bp in length  249335 249434: gap of unknown length  249435 250546: contig of 1112 bp in length.  Location/Qualifiers  1..250546  /organism="Rattus norvegicus"  /mol_type="genomic DNA"  /db_xref="taxon:10116"  /clone="CH230-6D11"  1..1827  /note="wgs end_extension  clone end:T7"  4183..5065  /note="clone_boundary  clone end:T7  site:EcoRI  end sequence:BH358077"  186843..227017  /note="clone_boundary  clone end:Sp6  site:EcoRI  end sequence:BH358081"  246532..249334  /note="wgs end_extension  clone end:Sp6"</p>			
FEATURES			
source			
misc_feature			
misc_feature			
misc_feature			
misc_feature			
ORIGIN			

```

Query Match          5.8%; Score 27; DB 2; Length 250546;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GCTGCTGCTGCTGCTGCTGCTTAA 80
Db 14787 GCTGCTGCTGCTGCTGCTGCTTAA 14813

RESULT 50
AC128474
LOCUS
DEFINITION
AC128474
ACCESSION
VERSION
GTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 258061)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,P.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgievski,S., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunnarson,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladuk,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulay,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensheva,L., Louisegh,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankweliemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Platter,F., Poinexter,A., Popovic,D., Primus,E., Pu.L.-L.,
Puzos,M., Quirroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,W., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission

```

TITLE

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 258061)  
Worley,K.C.

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 258061)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (21-SEP-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Sep 21, 2002 this sequence version replaced gi:21909191.

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the  
sequence may extend beyond the ends of the clone and there may be  
contigs that consist entirely of whole genome shotgun sequence  
reads. Both end sequences and whole genome shotgun sequence only  
contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GYUH

Center clone name: CH230-51A3

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 237914 bases at least Q40

Consensus quality: 240542 bases at least Q30

Consensus quality: 242364 bases at least Q20

Estimated insert size: 259476; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html))

\* NOTE: This sequence may represent more than one clone.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces

\* is believed to be correct as given, however the sizes

\* of the gaps between them are based on estimates that have

\* provided by the submittor.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.

\* 1 258061: contig of 258061 bp in length.

Location/Qualifiers

1. .258061

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-51A3"

1. .1429

/note="wgs end extension

clone\_end:T7"

8852. .9636

/note="clone boundary

clone\_end:T7"

site:EcoRI

end\_sequence:BH336930"

248189. .248838

/note="clone boundary

clone\_end:Sp6

site:EcoRI

end\_sequence:BH336932"

256785. .258061

/note="wgs end extension

clone\_end:Sp6"

ORIGIN

Query Match 5.8%; Score 27; DB 2; Length 258061;  
Best Local Similarity 100.0%; Pred. No. 0.0012;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 51 AACGCTGCTGCTGCTGCTGCTGCTGCT 77  
Db 82726 AACGCTGCTGCTGCTGCTGCTGCTGCT 82752

Search completed: May 25, 2004, 18:30:57  
Job time : 3357 secs